

# ON EXTREMALS FOR A RADON-LIKE TRANSFORM

MICHAEL CHRIST

ABSTRACT. The operator defined by convolution, in  $\mathbb{R}^d$ , with (affine) surface measure on a paraboloid satisfies a dilation-invariant  $L^p \rightarrow L^q$  inequality, and enjoys a high-dimensional group of symmetries. Extremal functions are shown to exist for this inequality. Moreover, any extremizing sequence is shown to be precompact modulo this symmetry group.

## 1. INTRODUCTION

If  $X, Y$  are Banach spaces, and if  $T : X \rightarrow Y$  is a bounded linear operator, then by an extremal vector for the associated inequality  $\|Tx\|_Y \leq C\|x\|_X$  is meant a nonzero vector  $x \in X$  such that  $\|Tx\|_Y = \|T\| \cdot \|x\|_X$ . The following are natural questions:

- (1) What is the norm of  $T$ ?
- (2) Does there exist at least one extremal vector  $x$ ?
- (3) Are extremal vectors unique, modulo scalar multiplication, or modulo explicitly known symmetries enjoyed by  $T$ ? If not, then what is the set of all extremal vectors?
- (4) If  $x$  is a *quasiextremal* vector in the sense that  $\|Tx\|_Y \geq c\|T\|\|x\|_X$  for a specified scalar  $c < 1$ , then what can be said quantitatively about the structure of  $x$ ? There are three natural regimes for this question: (a) for fixed  $c$ , (b) as  $c \rightarrow 0$ , and (c) as  $c \rightarrow 1$ .
- (5) What qualitative properties do extremal vectors enjoy?
- (6) One can likewise define extremal and quasiextremal pairs  $(x, y) \in X \times Y^*$ , where  $Y^*$  is the dual of  $Y$ ; an extremal pair is one which satisfies  $|y(Tx)| = \|T\| \|y\|_{Y^*} \|x\|_X$ . Each of questions (2) through (5) has an analogue for such pairs.

This paper investigates these questions, for a specific operator which arises in Euclidean harmonic analysis. This operator is modestly fundamental, as a basic example of a Fourier integral operator, of a Radon-like transform, and of an object whose analytic properties are governed by geometry and combinatorics. Within those frameworks it occupies a unique niche, attested to by an associated high-dimensional symmetry group, connected to the Heisenberg group and metaplectic representation.

Regard  $\mathbb{R}^d$  as  $\mathbb{R}^{d-1} \times \mathbb{R}$  with coordinates  $x = (x', x_d)$ . Let  $\sigma$  be the singular measure on  $\mathbb{R}^d$ , supported on the paraboloid  $\{x : x_d = |x'|^2\}$ , defined by  $\int f d\sigma =$

---

*Date:* April 16, 2009. Revised June 3, 2011.

The author was supported in part by NSF grants DMS-040126 and DMS-0901569.

$\int_{\mathbb{R}^{d-1}} f(x', |x'|^2) dx'$ .  $\sigma$  and surface measure on the paraboloid  $x_d = |x'|^2$  are mutually absolutely continuous, but  $\sigma$  enjoys a certain dilation symmetry which surface measure lacks. See §11 for discussion of a general context for such a measure.

The convolution operator and inequality under discussion are:

$$(1.1) \quad \mathbf{T}f = f * \sigma$$

$$(1.2) \quad \|\mathbf{T}f\|_{L^{d+1}(\mathbb{R}^d)} \leq \mathbf{A} \|f\|_{L^{(d+1)/d}(\mathbb{R}^d)}$$

for all  $f \in C^0 \cap L^\infty \cap L^1(\mathbb{R}^d)$ . Here  $\mathbf{A} < \infty$  denotes the optimal constant in the inequality, which depends only on the dimension  $d$ . For some discussion of the fundamental nature of this inequality, see [3].

In this paper we establish the existence of extremals, and of extremal pairs, for  $\mathbf{T}$  in all dimensions. Moreover, we obtain uniform quantitative information about the behavior of extremals, and of  $(1 - \delta)$ -quasiextremals with  $\delta$  close to 0. We show that sequences of approximate extremizers are precompact, after renormalization by an explicitly described symmetry group. This complements our earlier work [3] concerning  $c$ -quasiextremals, in the regime where  $c$  is not close to 1. Various natural questions remain open: We neither identify these extremals and the optimal constant  $\mathbf{A}$ , nor address the question of their uniqueness modulo symmetries,

One of our objectives is to begin to develop techniques which should be useful in an analysis of extremals for other closely related operators. See [5] for one such problem. This paper is the second in a series treating aspects of the meta-question: If the ratio  $\Phi(f) = \|\mathbf{T}f\|_q / \|f\|_p$  is large, then what are the properties of  $f$ ? The word “large” admits various interpretations. The initial work [3] is a study of those functions  $f$  for which the  $\Phi(f)$  is bounded below by some positive constant. In [6], qualitative properties of arbitrary critical points of  $\Phi$  are studied. The paper [4] demonstrates an equivalence between the inequality studied here, and a certain inequality for the Radon transform, and explicitly identifies all extremizers for both. This equivalence, together with an rearrangement inequality proved in [1] and special considerations for the subclass of all radially symmetric functions, could be used to give an alternative proof of the existence of extremizers for  $\Phi$ . However, the arguments given in the present paper, which ultimately rely on qualitative rather than exact symmetries, are more general and therefore retain some interest. One problem which has such qualitative but not exact symmetries is studied in [5].

We are indebted to René Quilodr  n and to Shuanglin Shao for useful advice on the exposition, and to Terence Tao and Shuanglin Shao for posing related questions to us.

## 2. RESULTS

Define

$$\mathbf{A} = \sup_{\|f\|_{(d+1)/d}=1} \|\mathbf{T}f\|_{d+1}.$$

**Definition 2.1.** An extremizer for the inequality (1.2) is a function  $f \in L^{(d+1)/d}(\mathbb{R}^d)$  which satisfies  $\|\mathbf{T}f\|_{L^{d+1}} = \mathbf{A} \|f\|_{L^{(d+1)/d}} \neq 0$ .

An extremizing sequence for the inequality (1.2) is a sequence of nonnegative functions  $f_\nu \in L^{(d+1)/d}$  satisfying

$$(2.1) \quad \|f_\nu\|_{L^{(d+1)/d}} \equiv 1$$

$$(2.2) \quad \|\mathbf{T}f_\nu\|_{L^{d+1}} \rightarrow \mathbf{A}.$$

For any  $\delta \in [0, 1)$ ,  $f \in L^{(d+1)/d}$  is a  $(1 - \delta)$ -quasiextremal for (1.2) if

$$(2.3) \quad \|\mathbf{T}f\|_{L^{d+1}} \geq (1 - \delta)\mathbf{A}\|f\|_{L^{(d+1)/d}} \neq 0.$$

In Definition 3.1 below we will introduce a group  $\mathcal{G}_d$  of diffeomorphisms of  $\mathbb{R}^d$  which are natural symmetries of our problem. Associated to each  $\phi \in \mathcal{G}_d$  is an invertible linear operator  $\phi^*$  on  $L^{(d+1)/d}(\mathbb{R}^d)$  satisfying  $\|\phi^*f\|_{(d+1)/d} = \|f\|_{(d+1)/d}$  and  $\|\mathbf{T}(\phi^*f)\|_{d+1} = \|\mathbf{T}(f)\|_{d+1}$  for all  $f \in L^{(d+1)/d}$ . Thus for any sequence of functions  $f_\nu$  and elements  $\phi_\nu \in \mathcal{G}_d$ ,  $(f_\nu)$  is an extremizing sequence if and only if  $(\phi_\nu^*f_\nu)$  is an extremizing sequence; extremizing sequences can only be characterized modulo the action of  $\mathcal{G}_d$ .

**Theorem 2.2.** (i) *There exist extremizers for the inequality (1.2).*

(ii) *Let  $\{f_\nu\}$  be any extremizing sequence for the inequality (1.2). Then there exist an extremal  $f$  for (1.2) satisfying  $\|f\|_{(d+1)/d} = 1$ , a subsequence  $\{f_{\nu_i}\}$ , and a sequence of symmetries  $\phi_i \in \mathcal{G}_d$  such that  $\phi_i^*(f_{\nu_i}) \rightarrow f$  in  $L^{(d+1)/d}$  norm.*

(iii) *There exist a constant  $C_0 < \infty$ , a function  $\Psi : [0, \infty) \rightarrow [0, \infty)$  satisfying  $\Psi(t) \geq t^{(d+1)/d}$  for all  $t$  and  $\Psi(t)/t^{(d+1)/d} \rightarrow \infty$  as  $t \rightarrow 0^+$  and also as  $t \rightarrow \infty$ , and a function  $\rho : [1, \infty) \rightarrow (0, \infty)$  satisfying  $\rho(R) \rightarrow 0$  as  $R \rightarrow \infty$ , with the following property. For any extremizing sequence  $\{f_\nu\}$  for the inequality (1.2), there exists a sequence of elements  $\phi_\nu \in \mathcal{G}_d$  such that for all sufficiently large  $\nu$ ,  $\phi_\nu^*f_\nu$  can be decomposed as  $g_\nu + h_\nu$  so that  $\|h_\nu\|_{(d+1)/d} \rightarrow 0$ ,*

$$(2.4) \quad \int_{\mathbb{R}^d} \Psi(g_\nu) \leq C_0;$$

for all  $R \geq 1$ ,

$$(2.5) \quad \int_{|x| \geq R} g_\nu(x)^{(d+1)/d} dx \leq \rho(R).$$

Moreover for any  $R \geq 1$ , for all  $\nu \geq \nu(R)$ ,  $g_\nu$  may be further decomposed as  $g_\nu^\sharp + g_\nu^\flat$  where both summands continue to satisfy (2.4) and (2.5), and moreover

$$(2.6) \quad \|g_\nu^\sharp\|_{C^1} \leq R \text{ and } \|g_\nu^\flat\|_{L^{(d+1)/d}} \leq \eta(R)$$

where  $\eta(R) \rightarrow 0$  as  $R \rightarrow \infty$ , and the function  $\eta$  is independent of the extremizing sequence.

(iv) *For any nonnegative extremizer  $f$  for (1.2) satisfying  $\|f\|_{(d+1)/d} = 1$ , there exists  $\phi \in \mathcal{G}_d$  such that  $\phi^*f$  satisfies (2.4), (2.5), and (2.6).*

(v) *Any complex-valued extremizer for (1.2) agrees almost everywhere with  $e^{i\theta}f$  for some  $C^\infty$  nonnegative function  $f$  and some constant  $\theta \in \mathbb{R}$ .*

(vi) Any nonnegative extremizer for inequality (1.2) with  $\|f\|_{(d+1)/d} = 1$  satisfies the Euler-Lagrange equation

$$(2.7) \quad \mathbf{T}^*([\mathbf{T}f]^d) = \mathbf{A}^{d+1} f^{1/d}$$

almost everywhere. For any compact set  $K \subset \mathbb{R}^d$  there exists  $c > 0$  such that  $f(x) \geq c$  for almost every  $x \in K$ .

In conclusion (vi),  $\mathbf{T}^*$  denotes the transpose of  $\mathbf{T}$ .

**Remark 2.3.** Extremizers of (1.2) do not belong to the Schwartz class. Indeed, if  $f$  is a continuous nonnegative solution of the Euler-Lagrange equation (2.7) which does not vanish identically, then  $\limsup_{|x| \rightarrow \infty} |x|^{d/2} f(x) > 0$ . For  $\mathbf{T}f$  is bounded away from zero in some ball  $B$ . The equation and nonnegativity together force  $f^{1/d} \geq c \mathbf{T}^*(\chi_B)$ , and consequently  $f(x', x_d) \geq c(1 + |x'|)^{-d}$  for all  $(x', x_d)$  belonging to a certain tubular neighborhood, of constant width, of a paraboloid. In such a region,  $|x'| \sim |x|^{1/2}$  as  $|x| \rightarrow \infty$ .

Quasiextremals for the inequality (1.2) were studied in [3]. These are by definition functions which satisfy  $\|\mathbf{T}f\|_{d+1} \geq \varepsilon \|f\|_{(d+1)/d}$  for an arbitrary constant  $\varepsilon > 0$ ;  $\varepsilon$  need not be close to the optimal constant  $\mathbf{A}$ . The analytic techniques introduced there were further developed in [10] to treat quasiextremals for the corresponding inequality for  $T_{\text{sphere}}$ . These techniques form the basis of the present paper. However, we have structured the exposition to emphasize certain geometric facets of the subject, especially the symmetry group and family of paraballs introduced in §3, which were less fully developed in [3].

Throughout the paper we assume that the dimension of the ambient space  $\mathbb{R}^d$  satisfies  $d \geq 2$ . Define

$$p = \frac{d+1}{d} \text{ and } q = d+1.$$

Since  $|T(f)| \leq T(|f|)$  and  $\|f\|_p = \|f\|_p$ , there exist extremal functions for the inequality in question if and only if there exist nonnegative extremal functions; we assume henceforth, without loss of generality and without further comment, that all functions under discussion are nonnegative.  $c, C, \gamma$  will denote finite positive constants which depend only on the dimension  $d$ , and which are permitted to change values from one occurrence to the next. Typically  $c$  will be small, while  $C$  will be large.

### 3. PRELIMINARIES

We begin by collecting various facts which will be used in the analysis.

**3.1. Symmetries.** The operator  $\mathbf{T}$  enjoys a relatively high-dimensional Lie group of symmetries. Let  $\Theta : \mathbb{R}^{d+d} \rightarrow \mathbb{R}$  be the function

$$\Theta(x, y) = x_d - y_d - |x' - y'|^2$$

and let  $\mathcal{I}$  be the incidence manifold  $\mathcal{I} = \{(x, y) \in \mathbb{R}^{d+d} : \Theta(x, y) = 0\}$ .  $\text{Diff}(\mathbb{R}^n)$  denotes the group of all  $C^\infty$  diffeomorphisms of  $\mathbb{R}^n$ .

**Definition 3.1.**  $\mathcal{G}_{d,d}$  denotes the set of all ordered pairs  $(\phi, \psi) \in \text{Diff}(\mathbb{R}^d) \times \text{Diff}(\mathbb{R}^d)$  which preserve  $\mathcal{I}$  in the strong sense that there exists  $0 \neq \lambda \in \mathbb{R}$  such that

$$(3.1) \quad \Theta(\phi(x), \psi(y)) = \lambda \Theta(x, y) \text{ for all } (x, y) \in \mathbb{R}^{d+d}.$$

$\mathcal{G}_d$  denotes the set of all  $\phi \in \text{Diff}(\mathbb{R}^d)$  for which there exists  $\psi$  such that  $(\phi, \psi) \in \mathcal{G}_{d,d}$ .

In particular,  $(x, y) \in \mathcal{I} \Rightarrow (\phi(x), \psi(y)) \in \mathcal{I}$ .

The following are examples of elements of  $\mathcal{G}_{d,d}$ :

- $(\phi(x), \psi(y)) = (x + v, y + v)$ , for any  $v \in \mathbb{R}^d$
- $(\phi(x), \psi(y)) = (rx', r^2x_d; ry', r^2y_d)$ , for any  $r \in \mathbb{R} \setminus \{0\}$
- $(\phi(x), \psi(y)) = (x' + u, x_d + 2u \cdot x' + |u|^2; y', y_d + 2u \cdot y')$ , for any  $u \in \mathbb{R}^{d-1}$
- $(\phi(x), \psi(y)) = (L(x'), x_d + |Lx'|^2 - |x'|^2; L^\dagger(y'), y_d - |L^\dagger(y')|^2 + |y'|^2)$ , for any invertible linear transformation  $L : \mathbb{R}^{d-1} \rightarrow \mathbb{R}^{d-1}$ , where  $L^\dagger$  is the inverse of the transpose of  $L$ .

Denote by  $\tau$  the involution  $x \mapsto -x$  of  $\mathbb{R}^d$ .

**Lemma 3.2.** (i) The correspondence  $(\phi, \psi) \leftrightarrow \phi$  is a bijection of  $\mathcal{G}_{d,d}$  with  $\mathcal{G}_d$ .

(ii) If  $(\phi, \psi) \in \mathcal{G}_{d,d}$  then  $\tau \circ \psi \circ \tau \in \mathcal{G}_d$ .

(iii) Every  $\phi \in \mathcal{G}_d$  is of the form  $(x', x_d) \mapsto (Lx', Ax_d + Q(x'))$  for some invertible affine endomorphism  $L$  of  $\mathbb{R}^{d-1}$ , some invertible affine endomorphism  $A$  of  $\mathbb{R}^1$ , and some polynomial  $Q : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$  of degree  $\leq 2$ . The pair  $(L, A)$  is unrestricted.  $Q$  takes the form  $Q = q + a$  where  $q$  is a homogeneous polynomial of degree 2 uniquely determined by  $(L, A)$ , while  $a$  is an arbitrary affine mapping.

*Proof.* (ii) is a direct consequence of the definitions.

To establish (iii), write  $\phi(x) = (F(x), f(x)) \in \mathbb{R}^{d-1} \times \mathbb{R}$  and similarly  $\psi(y) = (G(y), g(y))$ . To the equation

$$f(x) - g(y) - |F(x) - G(y)|^2 = \lambda(x_d - y_d - |x' - y'|^2)$$

apply any mixed second partial derivative  $\frac{\partial^2}{\partial x_i \partial y_j}$  to conclude that

$$\left\langle \frac{\partial F(x)}{\partial x_i}, \frac{\partial G(y)}{\partial y_j} \right\rangle \text{ is independent of } x, y.$$

Since  $\mathbb{R}^d \ni x \mapsto F(x) \in \mathbb{R}^{d-1}$  is a submersion at every  $x$ , it follows that  $\nabla G(y)$  is independent of  $y$ ; similarly  $\nabla F(x)$  is independent of  $x$ . Thus  $F, G$  are affine functions.

If  $i = d$  or  $j = d$  then  $\partial^2 / \partial x_i \partial y_j$  annihilates  $x_d - y_d - |x' - y'|^2$ , so it must annihilate  $F(x) \cdot G(y)$ . It follows again from the submersion property that  $F, G$  are independent of  $x_d, y_d$ .

By comparing terms we see that  $F(x) \cdot G(y) \equiv \lambda x' \cdot y'$  plus an affine function of  $x', y'$ . Therefore  $F, G$  take the form  $F(x) = Ax' + u$  and  $G(y) = \lambda By' + v$ , where  $B$  is the transpose of  $A^{-1}$  and  $u, v$  are vectors in  $\mathbb{R}^{d-1}$ .

Now  $f(x) - g(y)$  equals  $\lambda x_d - \lambda y_d$  plus a quadratic polynomial in  $x', y'$ . This forces  $f$  to be a quadratic polynomial in  $(x', x_d)$  in which the coefficient of  $x_d x_j$  vanishes for all  $1 \leq j \leq d$ . Likewise for  $g$ . The other parts of the description of  $\phi, \psi$  now follow from the equation.

In order to prove that  $(\phi, \psi) \rightarrow \phi$  is a bijection, which was conclusion (i), it suffices to prove that if  $\phi$  is the identity, then so must be  $\psi$ . We already know that  $\psi$  must take the form  $\psi(y', y_d) = (\lambda y' + v, g(y))$ . Thus

$$x_d - g(y) - |x' - y' - v|^2 \equiv \lambda(x_d - y_d - |x' - y'|^2).$$

Equating coefficients of  $x_d$  forces  $\lambda = 1$ . Therefore  $y_d \equiv g(y) + 2v \cdot (x' - y') + |v|^2$ . Since the left-hand side is independent of  $x'$ ,  $v$  must vanish, leaving  $g(y) \equiv y_d$ .  $\square$

Some further information concerning  $\mathcal{G}_d$  may be found in §10 below.

For each  $\phi \in \mathcal{G}_d$ , the Jacobian determinant  $J_\phi : \mathbb{R}^d \rightarrow (0, \infty)$  of  $\phi$  is a constant function. Associated to each element  $\phi \in \mathcal{G}_d$  is the mapping  $\phi^* : L^{(d+1)/d}(\mathbb{R}^d) \rightarrow L^{(d+1)/d}(\mathbb{R}^d)$  defined by  $\phi^* f(x) = f(\phi(x)) J^{d/(d+1)}$ . Then

$$(3.2) \quad \begin{aligned} \|\phi^* f\|_{(d+1)/d} &= \|f\|_{(d+1)/d} \\ \langle \psi^* g, \mathbf{T}(\phi^* f) \rangle &= \langle g, \mathbf{T}(f) \rangle \end{aligned}$$

for all  $f, g \in L^{(d+1)/d}$ .

### 3.2. Paraballs.

**Definition 3.3.** Let  $r_1, \dots, r_{d-1}, \rho > 0$ , let  $\mathbf{e} = \{e_1, \dots, e_{d-1}\}$  be any orthonormal basis for  $\mathbb{R}^{d-1}$ , let  $\mathbf{r} = (r_1, \dots, r_{d-1}) \in (0, \infty)^{d-1}$ , and let  $z = (\bar{x}, \bar{x}_*) \in \mathcal{I}$ , that is,  $z \in \mathbb{R}^d \times \mathbb{R}^d$  satisfies  $(\bar{x}_*)_d - \bar{x}_d = |\bar{x}'_* - \bar{x}'|^2$ . The *paraball*  $B = B(z, \mathbf{e}, \mathbf{r}, \rho)$  associated to these data is the set of all  $x \in \mathbb{R}^d$  satisfying

$$(3.3) \quad \sum_{j=1}^{d-1} r_j^{-2} |\langle x' - \bar{x}', e_j \rangle|^2 < 1,$$

$$(3.4) \quad |x_d - (\bar{x}_*)_d - |x' - \bar{x}'_*|^2| < \rho.$$

Here  $\bar{x} = (\bar{x}', \bar{x}_d) \in \mathbb{R}^d$  and  $\bar{x}^* = (\bar{x}'_*, (\bar{x}_*)_d) \in \mathbb{R}^d$ .

For any  $\lambda \geq 1$ , the expanded paraball  $\lambda B(z, \mathbf{e}, \mathbf{r}, \rho)$  is defined to be the set of all  $x \in \mathbb{R}^d$  satisfying

$$(3.5) \quad \sum_{j=1}^{d-1} r_j^{-2} |\langle x' - \bar{x}', e_j \rangle|^2 < \lambda^2,$$

$$(3.6) \quad |x_d - (\bar{x}_*)_d - |x' - \bar{x}'_*|^2| < \lambda \rho.$$

The dual paraball  $B_* = B_*(z, \mathbf{e}, \mathbf{r}, \rho)$  associated to these data is the set of all  $x^* = (x'_*, (x_*)_d) \in \mathbb{R}^d$  satisfying

$$(3.7) \quad \sum_{j=1}^{d-1} (r_j^*)^{-2} |\langle x'_* - \bar{x}'_*, e_j \rangle|^2 < 1$$

$$(3.8) \quad |(x_*)_d - \bar{x}_d + |x'_* - \bar{x}'_*|^2| < \rho$$

where  $r_j r_j^* = \rho$  for every  $j \in \{1, 2, \dots, d-1\}$ .

$\mathcal{B}(z, \mathbf{e}, \mathbf{r}, \rho)$  denotes the ordered pair  $\mathcal{B} = (B(z, \mathbf{e}, \mathbf{r}, \rho), B_*(z, \mathbf{e}, \mathbf{r}, \rho))$ .

$B$  uniquely determines  $B_\star$ , and vice versa. The point  $z = (\bar{x}, \bar{x}_\star)$  is an element of  $\mathbb{R}^d \times \mathbb{R}^d$ , not of  $\mathbb{R}^d$ . The data  $(\bar{x}, \mathbf{e}, \mathbf{r}, \rho)$  do not suffice to completely determine  $B$ ; the point  $\bar{x}$  can be regarded as a “center” of  $B$ , but the geometry of  $B$  depends also on  $\bar{x}_\star$ .

The mapping from data  $(z, \mathbf{e}, \mathbf{r}, \rho)$  to paraballs is not a one-to-one correspondence;  $O(d-1)$  acts naturally on the set of all orthonormal bases  $\mathbf{e}$ , and if  $\mathbf{e}, \tilde{\mathbf{e}}$  belong to the same orbit under this action then  $B(z, \mathbf{e}, \mathbf{r}, \rho) = B(z, \tilde{\mathbf{e}}, \mathbf{r}, \rho)$ .

Given  $\phi \in \mathcal{G}_d$  and a parabol  $B$ ,  $\{x : \phi(x) \in B\}$  is a parabol. For  $\phi$  can be expressed in the form  $\phi(x', x_d) = v + (L(x'), \lambda x_d + Q(x'))$  where  $L \in Gl(d-1)$ ,  $0 \neq \lambda \in \mathbb{R}$ ,  $v \in \mathbb{R}^d$ , and  $Q : \mathbb{R}^{d-1} \rightarrow \mathbb{R}^1$ . Then  $L^*L$  is a positive definite symmetric matrix, hence admits a factorization  $L^*L = O^{-1}DO$  with  $O \in O(d-1)$  and  $D$  a diagonal matrix with positive diagonal entries. The condition  $|L(x')|^2 \leq 1$  is equivalent to  $|\langle D^{1/2}O(x'), D^{1/2}O(x') \rangle| \leq 1$ , which can be expressed in terms of  $\mathbf{e}, \mathbf{r}$  where  $\mathbf{e}$  is the image of the standard basis for  $\mathbb{R}^{d-1}$  under  $O$ , and  $(r_1, \dots, r_{d-1})$  are the diagonal entries of  $D^{1/2}$ . Thus (3.3) transforms under  $\phi$  to another inequality of the same form. This is the main step in establishing the first conclusion of the next lemma.

**Lemma 3.4.** *There are natural actions of  $\mathcal{G}_d$  on the set of all paraballs, and of  $\mathcal{G}_{d,d}$  on the set of all dual pairs of paraballs. These actions are transitive.*

The simple remainder of the verification is left to the reader.

**3.3. Distance between paraballs.** It will be useful to quantify the notion that two paraballs are far apart. An approximate measure  $\varrho(B^\sharp, B^\flat)$  of the discrepancy between two paraballs is defined as follows.

To any parabol  $B = B(\bar{x}, \mathbf{e}, \mathbf{r}, \rho)$  is associated the balanced convex subset  $\mathcal{C} \subset \mathbb{R}^{d-1}$  defined by

$$\mathcal{C} = \left\{ v \in \mathbb{R}^{d-1} : \sum_{j=1}^{d-1} r_j^{-2} |\langle v, e_j \rangle|^2 < 1 \right\}.$$

Much of the geometry of  $B$  is encoded by  $\mathcal{C}$ ;  $B$  is the set of all  $(x', x_d)$  such that  $x' - \bar{x}' \in \mathcal{C}$  and  $|x_d - h(x')| < \rho$ , for a certain quadratic polynomial  $h$  specified by  $\bar{x}_\star$ .

In the following definition,  $\mathcal{C}^\sharp, \mathcal{C}^\flat$  denote the convex sets associated to  $B^\sharp, B^\flat$ , respectively.

**Definition 3.5.** Let  $B^\sharp = B(z^\sharp, \mathbf{e}^\sharp, \mathbf{r}^\sharp, \rho^\sharp)$  and  $B^b = B(z^b, \mathbf{e}^b, \mathbf{r}^b, \rho^b)$  be two paraballs, where  $z^\sharp = (\bar{x}^\sharp, \bar{x}_\star^\sharp)$ ,  $z^b = (\bar{x}^b, \bar{x}_\star^b)$  are two points in  $\mathcal{I}$ . Define

$$\begin{aligned} \varrho(B^\sharp, B^b) = & \frac{\max(\rho^\sharp, \rho^b)}{\min(\rho^\sharp, \rho^b)} + \sup_{v \in \mathcal{C}^\sharp} \sum_{j=1}^{d-1} \frac{|\langle v, e_j^b \rangle|^2}{(r_j^b)^2} + \sup_{v \in \mathcal{C}^b} \sum_{j=1}^{d-1} \frac{|\langle v, e_j^\sharp \rangle|^2}{(r_j^\sharp)^2} \\ & + \sum_{j=1}^{d-1} \frac{|\langle \bar{x}^{\sharp'} - \bar{x}^{b'}, e_j^\sharp \rangle|^2}{(r_j^\sharp)^2} + \sum_{j=1}^{d-1} \frac{|\langle \bar{x}^{\sharp'} - \bar{x}^{b'}, e_j^b \rangle|^2}{(r_j^b)^2} \\ & + \sum_{j=1}^{d-1} \frac{|\langle \bar{x}_\star^{\sharp'} - \bar{x}_\star^{b'}, e_j^\sharp \rangle|^2}{(\rho^\sharp/r_j^\sharp)^2} + \sum_{j=1}^{d-1} \frac{|\langle \bar{x}_\star^{\sharp'} - \bar{x}_\star^{b'}, e_j^b \rangle|^2}{(\rho^b/r_j^b)^2} \\ & + \frac{|\bar{x}_d^b - (\bar{x}_\star^\sharp)_d - |\bar{x}^{b'} - \bar{x}_\star^{\sharp'}||}{\rho^\sharp} + \frac{|\bar{x}_d^\sharp - (\bar{x}_\star^b)_d - |\bar{x}^{\sharp'} - \bar{x}_\star^{b'}||}{\rho^b}. \end{aligned}$$

In particular,  $\varrho(B^\sharp, B^b) \geq 1$  and  $\varrho(B^\sharp, B^b) = \varrho(B^b, B^\sharp)$  for any  $B^\sharp, B^b$ , and  $\varrho(B, B) = 1$  for any  $B$ .

$\varrho$  is one of many ways in which the difference between two paraballs can be quantified, and other variants would serve our purpose equally well. Of primary interest here is the situation in which two paraballs differ markedly, rather than when they nearly coincide.  $\varrho$  is not actually a metric; for instance,  $\varrho(B^\sharp, B^b) \geq 1$  for any  $B^\sharp, B^b$ .

The following two invariance properties are direct consequences of Definitions 3.3 and 3.5.

**Lemma 3.6.** *For any  $\phi \in \mathcal{G}_d$  and any two paraballs  $B^\sharp, B^b$ ,*

$$(3.9) \quad \varrho(\phi(B^\sharp), \phi(B^b)) = \varrho(B^\sharp, B^b).$$

*For any two dual pairs of paraballs  $(B^\sharp, B_\star^\sharp)$  and  $(B^b, B_\star^b)$ ,*

$$(3.10) \quad \varrho(B^\sharp, B^b) = \varrho(B_\star^\sharp, B_\star^b).$$

The next lemma relates the distance between two paraballs to the relative size of their intersection.

**Lemma 3.7.** *There exists a constant  $C < \infty$  which depends only on the dimension  $d$ , such that for any two paraballs  $B^\sharp, B^b$ ,*

$$(3.11) \quad \varrho(B^\sharp, B^b) \leq C \left( \frac{\max(|B^\sharp|, |B^b|)}{|B^\sharp \cap B^b|} \right)^C.$$

There is of course no converse inequality; the right-hand side becomes infinite whenever  $B^\sharp \cap B^b = \emptyset$ .

The following elementary fact will be used in the proof: For any  $d \geq 1$  there exist constants  $C, c \in (0, \infty)$  such that for any convex set  $\mathcal{C} \subset \mathbb{R}^d$  of positive Lebesgue measure, and any quadratic polynomial  $Q : \mathbb{R}^d \rightarrow \mathbb{R}$  which does not vanish identically, for any  $\varepsilon > 0$

$$|\{y \in \mathcal{C} : |Q(y)| < \varepsilon \sup_{\mathcal{C}} |Q|\}| \leq C \varepsilon^c |\mathcal{C}|.$$



*Proof of Lemma 3.7.* We need to show that a lower bound on  $\varrho(B^\sharp, B^\flat)$  implies an upper bound on  $|B^\sharp \cap B^\flat| / \max(|B^\sharp|, |B^\flat|)$  of inverse power law type. We may assume that  $\varrho(B^\sharp, B^\flat)$  is large, since otherwise the inequality holds automatically for large enough  $C$ .

Recall the description of  $B^\flat$  as the set of all  $(x', x_d)$  such that  $x' \in \mathcal{C}^\flat$ , and  $|x_d - h(x')| < \rho$ , for a certain function  $h$  which depends on  $B^\flat$ . This implies that

$$\begin{aligned} |B^\sharp \cap B^\flat| &\leq \min(\rho^\sharp, \rho^\flat) |\mathcal{C}^\sharp \cap \mathcal{C}^\flat|. \\ &\leq \frac{\min(\rho^\sharp, \rho^\flat)}{\max(\rho^\sharp, \rho^\flat)} \max(|B^\sharp|, |B^\flat|) \end{aligned}$$

If  $\frac{\max(\rho^\sharp, \rho^\flat)}{\min(\rho^\sharp, \rho^\flat)} \gtrsim \varrho(B^\sharp, B^\flat)$ , this concludes the proof. In the same way,

$$\begin{aligned} |B^\sharp \cap B^\flat| &\leq \frac{|(\bar{x}^\sharp + \mathcal{C}^\sharp) \cap (\bar{x}^\flat + \mathcal{C}^\flat)|}{\max(|\mathcal{C}^\sharp|, |\mathcal{C}^\flat|)} \max(|B^\sharp|, |B^\flat|). \\ &\leq \frac{\min(|\mathcal{C}^\sharp|, |\mathcal{C}^\flat|)}{\max(|\mathcal{C}^\sharp|, |\mathcal{C}^\flat|)} \max(|B^\sharp|, |B^\flat|). \end{aligned}$$

The quantity  $\frac{\max(|\mathcal{C}^\sharp|, |\mathcal{C}^\flat|)}{\min(|\mathcal{C}^\sharp|, |\mathcal{C}^\flat|)}$  is comparable to the sum of the second and third terms in the definition of  $\varrho(B^\sharp, B^\flat)$ . The desired inequality follows, if either of these terms is  $\gtrsim \varrho(B^\sharp, B^\flat)$ . Moreover

$$|(\bar{x}^\sharp + \mathcal{C}^\sharp) \cap (\bar{x}^\flat + \mathcal{C}^\flat)| \leq C \left( \sum_{j=1}^{d-1} |\langle \bar{x}^{\sharp'} - \bar{x}^{\flat'}, e_j^\sharp \rangle| / r_j^\sharp + \sum_{j=1}^{d-1} |\langle \bar{x}^{\sharp'} - \bar{x}^{\flat'}, e_j^\flat \rangle| / r_j^\flat \right) \max(|\mathcal{C}^\sharp|, |\mathcal{C}^\flat|),$$

so the conclusion holds if either the fourth or fifth term in the definition of  $\varrho$  is largest.

Consider next the final two terms in the definition. Define  $Q^\sharp(y) = y_d - (\bar{x}_\star^\sharp)_d - |y' - \bar{x}_\star^{\sharp'}|^2$  and  $Q^\flat(y) = y_d - (\bar{x}_\star^\flat)_d - |y' - \bar{x}_\star^{\flat'}|^2$ . Consider the case where the eighth term is large, in the sense that  $|Q^\sharp(\bar{x}^\flat)| \geq \frac{1}{2} \varrho(B^\sharp, B^\flat) \rho^\sharp$ , and the first seven terms in the definition of  $\varrho$  are all  $\leq c_0 \varrho(B^\sharp, B^\flat)$  for a suitably small constant  $c_0$ , and moreover,

$$\max(\rho^\sharp, \rho^\flat) / \min(\rho^\sharp, \rho^\flat) \leq \varrho(B^\sharp, B^\flat)^{1/2}.$$

Then  $\bar{x}^\flat \notin B^\sharp$ . We aim to prove that  $|B^\sharp \cap B^\flat|$  is relatively small. To this end consider the quadratic polynomial  $P : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$  defined by  $P(z) = Q^\sharp(z, t(z))$  where  $t(z)$  is chosen so that  $Q^\flat(z, t(z)) \equiv 0$ , that is,  $t(z) = (\bar{x}_\star^\flat)_d + |z - \bar{x}_\star^{\flat'}|^2$ . Observe that

$$\text{If } |P(z)| > \rho^\flat + \rho^\sharp \text{ then } B^\sharp \cap B^\flat \cap (\{z\} \times \mathbb{R}) = \emptyset.$$

In the case which we are now analyzing,  $P$  satisfies

$$|P(\bar{x}^\flat)| \geq \frac{1}{2} \varrho(B^\sharp, B^\flat) \rho^\sharp.$$

Define

$$\varepsilon = \frac{3 \max(\rho^\sharp, \rho^\flat)}{\varrho(B^\sharp, B^\flat) \rho^\sharp} \leq 3 \varrho(B^\sharp, B^\flat)^{-1/2}.$$

On  $\mathcal{C}^\flat$ ,

$$|P(z)| \geq \varepsilon \varrho(B^\sharp, B^\flat) \rho^\sharp = 3 \max(\rho^\sharp, \rho^\flat) > \rho^\sharp + \rho^\flat,$$

for all  $y$  outside a set of measure  $\leq C\varepsilon^c|\mathcal{C}^b|$ .

We conclude that

$$\begin{aligned} |B^\sharp \cap B^\flat| &\leq C\varepsilon^c \min(\rho^\sharp, \rho^\flat) |\mathcal{C}^b| \\ &\leq C' \varrho(B^\sharp, B^\flat)^{-c'} |B^\flat| \end{aligned}$$

for some  $c', C' \in \mathbb{R}^+$ . This is the bound required.

Everything is symmetric in the indices  $\sharp, \flat$ , so the case where the ninth term is large requires no further discussion. Likewise it suffices now to treat the case where the sixth term is large, the seventh term being handled by symmetry.

Suppose then that  $|\langle \bar{x}_\star^\sharp - \bar{x}_\star^\flat, e_i^\sharp \rangle| \geq c\varrho(B^\sharp, B^\flat)\rho^\sharp/r_i$  for some index  $i$ . To simplify notation write  $u = \bar{x}_\star^\sharp$  and  $v = \bar{x}_\star^\flat$ . If  $y = (w, t) \in \mathbb{R}^{d-1} \times \mathbb{R}$  belongs to  $B^\sharp \cap B^\flat$  then  $t$  satisfies  $|t - (\bar{x}_\star^\sharp)_d - |w - u|^2| < \rho^\sharp$  and  $|t - (\bar{x}_\star^\flat)_d - |w - v|^2| < \rho^\flat$ . Subtracting gives

$$|2\langle w, u - v \rangle - s| < 2\max(\rho^\sharp, \rho^\flat)$$

where  $s = 2|u|^2 - 2u \cdot v$ . The value of  $s$  is little consequence. With respect to the basis  $\mathbf{e}^\sharp$ , the  $i$ -th component of  $u - v$  has absolute value  $\geq c\varrho(B^\sharp, B^\flat)\rho^\sharp/r_i$ .  $w$  ranges over a translate of  $\mathcal{C}^\sharp$ ; in particular,  $\langle w, \mathbf{e}_i^\sharp \rangle$  satisfies the sole constraint  $\langle w, \mathbf{e}_i^\sharp \rangle < r_i$ . This forces

$$|\{w \in \bar{x}_\star^\sharp + \mathcal{C}^\sharp : |2\langle w, u - v \rangle - s| < 2\max(\rho^\sharp, \rho^\flat)\}| \leq C\varrho(B^\sharp, B^\flat)^{-1}|\mathcal{C}^\sharp|,$$

uniformly in all  $s \in \mathbb{R}$ . This implies the required bound, by a repetition of arguments given above.  $\square$

There is an adequate quasi-triangle inequality for  $\varrho$ .

**Lemma 3.8.** *For any three paraballs,*

$$(3.12) \quad \varrho(B^\sharp, B^\flat) \leq C\varrho(B^\sharp, B^\natural)^C + C\varrho(B^\natural, B^\flat)^C$$

where  $C < \infty$  depends only on the dimension  $d$ .

By exploiting the transitive action of  $\mathcal{G}_d$ , we may assume without loss of generality that  $B^\natural = B(z, \mathbf{e}, \mathbf{r}, \rho)$  where  $z = (0, 0)$ ,  $\mathbf{e}$  is the standard basis for  $\mathbb{R}^{d-1}$ ,  $\mathbf{r} = (1, 1, \dots, 1)$  and  $\rho = 1$ . Reasoning like that in the proof of Lemma 3.7 then controls the parameters specifying  $B^\sharp$  in terms of  $\eta^{-1} = \varrho(B^\sharp, B^\natural)$ , and likewise for  $B^\flat$ . It then follows directly from the definitions that  $\varrho(B^\sharp, B^\flat) \leq C\eta^{-C}$  where now  $\eta$  is the maximum of the two quasidistances.

### 3.4. Lorentz spaces.

**Definition 3.9.** Let  $f$  be a nonnegative function which is finite almost everywhere. A rough level set decomposition of  $f$  is a representation  $f = \sum_{j=-\infty}^{\infty} 2^j f_j$  where  $f_j$  is supported on a set  $E_j$ ,  $1 \leq |f_j(x)| < 2$  for almost every  $x$ , and the sets  $E_j$  are pairwise disjoint and measurable.

Any (complex-valued) function  $f$  which is finite almost everywhere on  $\mathbb{R}^d$  admits such a decomposition, which is unique modulo redefinition on sets of Lebesgue measure zero. As shorthand for such a decomposition we will write “ $f = \sum_j 2^j f_j$ ,  $f_j \leftrightarrow E_j$ ”.

A quasinorm for the Lorentz space  $L^{p,r}(\mathbb{R}^d)$  is

$$\|f\|_{p,r} = \left( \sum_j (2^j |E_j|^{1/p})^r \right)^{1/r},$$

with the natural interpretation for  $r = \infty$ . For  $p > 1$  and  $r \in [1, \infty]$  there exist equivalent expressions which satisfy the triangle inequality. These spaces are nested:  $L^{p,r} \subset L^{p,s}$  if  $r \leq s$ , and the inclusion is proper if  $r < s$ . See [12] for further information.

**3.5. Analytic preliminaries.** We review here four facts established in [3]. The first of these results asserts that any quasiextremal function for inequality (1.2) is closely connected with the characteristic function of some parabol of comparable  $L^p$  norm.

**Lemma 3.10.** [3] *For any  $\varepsilon > 0$  there exist  $c, C \in \mathbb{R}^+$  with the following property. If  $f \in L^p$  is a nonnegative function with rough level set decomposition  $f = \sum_{j \in \mathbb{Z}} 2^j f_j$ ,  $f_j \leftrightarrow E_j$ , and if  $\|\mathbf{T}f\|_q \geq \varepsilon \|f\|_p$  then there exist an index  $j$  and a parabol  $B$  such that*

$$\|2^j f_j \cdot \chi_B\|_p \geq c\varepsilon^C \|f\|_p$$

and

$$|B| \leq |E_j|.$$

This is Theorem 1.5 of [3].

It is often useful to work with the bilinear form  $\langle g, \mathbf{T}f \rangle$ . The following result connects quasiextremal pairs  $(f, g)$  with dual pairs of paraballs, in the basic case when  $f, g$  are both characteristic functions of sets.

**Proposition 3.11.** [3] *There exist positive finite constants  $C, c, \gamma$ , depending only on the dimension  $d$ , with the following property. Let  $E, E^*$  be measurable subsets of  $\mathbb{R}^d$  satisfying  $0 < |E|, |E^*| < \infty$ . Define  $\varepsilon > 0$  to be*

$$\varepsilon = \frac{\langle \chi_{E^*}, \mathbf{T}\chi_E \rangle}{|E|^{d/(d+1)} |E^*|^{d/(d+1)}}.$$

*Then there exists a pair  $\mathcal{B} = (B, B^*)$  of dual paraballs such that*

$$\begin{aligned} |B| &\leq |E|, & |E \cap B| &\geq c\varepsilon^\gamma |E|, \\ |B^*| &\leq |E^*|, & |E^* \cap B^*| &\geq c\varepsilon^\gamma |E^*|. \end{aligned}$$

Under these hypotheses, Theorem 1.2 of [3] gives  $|B| \leq |E|$ ,  $|B^*| \leq |E^*|$ , and

$$\langle \mathbf{T}(\chi_{E \cap B}), \chi_{E^* \cap B^*} \rangle \geq c\varepsilon^C \langle \mathbf{T}(\chi_E), \chi_{E^*} \rangle = c\varepsilon^{C+1} |E|^{d/(d+1)} |E^*|^{d/(d+1)}.$$

Since  $\langle \mathbf{T}(\chi_{E \cap B}), \chi_{E^* \cap B^*} \rangle \leq \mathbf{A} |E \cap B|^{d/(d+1)} |E^* \cap B^*|^{d/(d+1)}$ , the lower bounds  $|E \cap B| \geq c\varepsilon^C |E|$  and  $|E^* \cap B^*| \geq c\varepsilon^C |E^*|$  stated in Proposition 3.11 follow directly.

Although the inequality  $\|\mathbf{T}f\|_q \lesssim \|f\|_p$  is dilation-invariant and is the only  $L^p(\mathbb{R}^d) \rightarrow L^q(\mathbb{R}^d)$  inequality valid for  $\mathbf{T}$ , it is nonetheless a suboptimal inequality within the more general context of Lorentz spaces. The following is Theorem 1.6 of [3].

**Proposition 3.12.**  $\mathbf{T}$  maps  $L^{p,r}$  boundedly to  $L^q$  for all  $p < r < q$ .

The following was a principal ingredient in the proof of Proposition 3.12, and will be needed again below. See Lemma 9.2 and inequality (9.10) of [3].

**Lemma 3.13.** *For any  $d \geq 2$  there exist  $C, C' < \infty$  with the following property. Let  $E, E', F \subset \mathbb{R}^d$  be measurable sets with positive, finite measures. Let  $\eta \in (0, \mathbf{A}]$ . If  $T\chi_E(x) \geq \eta|E|^{1/p}|F|^{-1+1/p}$  and  $T\chi_{E'}(x) \geq \eta|E'|^{1/p}|F|^{-1+1/p}$  for every  $x \in F$ , then  $|E'| \leq C'\eta^{-C}|E|$ .*

The exponents in the hypotheses are natural; the hypotheses imply for instance that  $\langle \chi_F, T\chi_E \rangle \geq \eta|F|^{1/p}|E|^{1/p}$ .

#### 4. DISTANT PARABALLS INTERACT WEAKLY

The following lemma may at present seem unmotivated, but will later provide, in the proof of Lemma 7.2, the geometric input for perhaps the most central step of our analysis. By a partition of a set we will always mean an expression as a union of pairwise disjoint subsets.

**Lemma 4.1.** *For each  $d \geq 2$  there exists  $C < \infty$  with the following property. Let  $\eta \in (0, 1]$ . Let  $\{B_\alpha : \alpha \in S\}$  be an arbitrary finite collection of paraballs in  $\mathbb{R}^d$  satisfying  $\varrho(B_\alpha, B_\beta) \geq C\eta^{-C}$  whenever  $\alpha \neq \beta$ . Let  $F \subset \mathbb{R}^d$  be a Lebesgue measurable set of finite measure. Then  $F$  can be measurably partitioned as  $F = \cup_{\alpha \in S} F_\alpha$  in such a way that*

$$(4.1) \quad \langle \chi_{F_\beta}, \mathbf{T}\chi_{B_\alpha} \rangle \leq \eta|F|^{1/p}|B_\alpha|^{1/p} \text{ whenever } \alpha \neq \beta.$$

Here  $C$  depends only on the dimension  $d$ , not on  $\eta$ .

*Proof.* Define

$$\gamma_\beta = \frac{1}{3}\eta|F|^{-1+1/p}|B_\beta|^{1/p}$$

and

$$\tilde{F}_\beta = \{x \in F : \mathbf{T}\chi_{B_\beta}(x) > \gamma_\beta\},$$

noting that

$$(4.2) \quad \langle \chi_{F \setminus \tilde{F}_\alpha}, \mathbf{T}\chi_{B_\alpha} \rangle \leq \int_{F \setminus \tilde{F}_\alpha} \gamma_\alpha \leq \gamma_\alpha|F| = \frac{1}{3}\eta|F|^{1/p}|B_\alpha|^{1/p}.$$

Choose pairwise disjoint measurable sets  $F_\beta \subset \tilde{F}_\beta$  so that  $\cup_\beta \tilde{F}_\beta = \cup_\beta F_\beta$ . Their union is not necessarily all of  $F$ , but  $F^\dagger = F \setminus \cup_\beta F_\beta$  already satisfies  $\langle \chi_{F^\dagger}, \mathbf{T}\chi_{B_\alpha} \rangle \leq \frac{1}{3}\eta|F|^{1/p}|B_\alpha|^{1/p}$  for every  $\alpha$  by (4.2), so it suffices to prove that for all  $\alpha \neq \beta$ ,

$$(4.3) \quad \langle \chi_{F_\beta}, \mathbf{T}\chi_{B_\alpha} \rangle \leq \frac{2}{3}\eta|F|^{1/p}|B_\alpha|^{1/p}.$$

We prove (4.3) by contradiction. Suppose that there exist indices  $\alpha \neq \beta \in S$  for which (4.3) fails to hold. These indices will remain fixed for the remainder of this proof. We aim to prove that  $\varrho(B_\alpha, B_\beta)$  is small, contradicting the hypothesis.

Set  $\mathcal{F} = F_\beta \cap \tilde{F}_\alpha$ . Then  $\langle \chi_{F_\beta \setminus \tilde{F}_\alpha}, \mathbf{T}\chi_{B_\alpha} \rangle \leq \frac{1}{3}\eta|F_\beta|^{1/p}|B_\alpha|^{1/p}$  as in (4.2), so

$$(4.4) \quad \langle \chi_{\mathcal{F}}, \mathbf{T}\chi_{B_\alpha} \rangle = \langle (\chi_{F_\beta} - \chi_{F_\beta \setminus \tilde{F}_\alpha}), \mathbf{T}\chi_{B_\alpha} \rangle \geq \frac{1}{3}\eta|F|^{1/p}|B_\alpha|^{1/p}.$$

Since  $\langle \chi_{\mathcal{F}}, \mathbf{T}\chi_{B_\alpha} \rangle \leq \mathbf{A}|\mathcal{F}|^{1/p}|B_\alpha|^{1/p}$  by definition of  $\mathbf{A}$ , this forces

$$|\mathcal{F}| \geq 3^{-p}\eta^p \mathbf{A}^{-p}|F|.$$

Given  $B_\alpha, B_\beta, \mathcal{F}$  as above, apply Proposition 3.11 with  $E = B_\alpha$ ,  $E^* = \mathcal{F}$  to obtain a pair  $\mathcal{B}^\alpha(z_\alpha, \mathbf{e}_\alpha, \mathbf{r}_\alpha, \rho_\alpha) = (B^\alpha, B_\star^\alpha)$  satisfying

$$\begin{aligned} |B^\alpha| &\leq |B_\alpha|, & |B_\star^\alpha| &\leq |\mathcal{F}| \leq |F|, \\ |B^\alpha \cap B_\alpha| &\geq c\eta^\gamma |B_\alpha|, & |B_\star^\alpha \cap \mathcal{F}| &\geq c\eta^\gamma |\mathcal{F}| \geq c\eta^\gamma |F| \end{aligned}$$

for certain constants  $c, \gamma$ , whose values have changed from one occurrence to the next.

Set  $\tilde{\mathcal{F}} = \mathcal{F} \cap B_\star^\alpha$ . We know already that  $|\tilde{\mathcal{F}}| \geq c\eta^\gamma |F|$ . Since  $\mathbf{T}\chi_{B_\beta}(x) > \gamma_\beta$  for every  $x \in F_\beta \supset \mathcal{F} \supset \tilde{\mathcal{F}}$ ,

$$\langle \chi_{\tilde{\mathcal{F}}}, \mathbf{T}\chi_{B_\beta} \rangle \geq \gamma_\beta |\tilde{\mathcal{F}}| = \frac{1}{3}\eta |\tilde{\mathcal{F}}| \cdot |F|^{-1+1/p} |B_\beta|^{1/p} \geq c\eta^\gamma |\tilde{\mathcal{F}}|^{1/p} |B_\beta|^{1/p}$$

for certain positive constants  $c, \gamma$ . Consequently Proposition 3.11 can be applied again, this time with  $E = B_\beta$  and  $E^* = \tilde{\mathcal{F}}$ , to obtain a pair  $\mathcal{B}^\beta(z_\beta, \mathbf{e}_\beta, \mathbf{r}_\beta, \rho_\beta) = (B^\beta, B_\star^\beta)$  satisfying

$$\begin{aligned} |B^\beta| &\leq |B_\beta|, & |B_\star^\beta| &\leq |\tilde{\mathcal{F}}| \leq |F|, \\ |B^\beta \cap B_\beta| &\geq c\eta^\gamma |B_\beta|, & |B_\star^\beta \cap \tilde{\mathcal{F}}| &\geq c\eta^\gamma |\tilde{\mathcal{F}}| \geq c\eta^\gamma |F| \end{aligned}$$

for certain constants  $c, \gamma$ .

Since  $B_\star^\beta \cap B_\star^\alpha \supset B_\star^\beta \cap B_\star^\alpha \cap \mathcal{F} = B_\star^\beta \cap \tilde{\mathcal{F}}$ ,

$$|B_\star^\beta \cap B_\star^\alpha| \geq |B_\star^\beta \cap \tilde{\mathcal{F}}| \geq c\eta^\gamma |F| \geq c\eta^\gamma \max(|B_\star^\beta|, |B_\star^\alpha|).$$

By Lemmas 3.7 and 3.6, this implies that

$$\varrho(B^\alpha, B^\beta) \leq C\eta^{-C}.$$

Since  $|B^\alpha| \leq |B_\alpha|$  and  $|B^\beta| \leq |B_\beta|$ , while  $|B^\alpha \cap B_\alpha| \geq c\eta^\gamma |B_\alpha|$  and  $|B^\beta \cap B_\beta| \geq c\eta^\gamma |B_\beta|$ , one has

$$\varrho(B^\beta, B_\beta) \leq C\eta^{-C} \text{ and } \varrho(B^\alpha, B_\alpha) \leq C\eta^{-C}$$

by Lemma 3.7. Therefore by the quasi-triangle inequality of Lemma 3.8,

$$\varrho(B_\alpha, B_\beta) \leq C\eta^{-C}.$$

This contradicts the assumption that  $\varrho(B_\alpha, B_\beta)$  is sufficiently large.  $\square$

## 5. STEP 1: ENTROPY REFINEMENT

According to Proposition 3.12, the inequality (1.2), while scale-invariant, is not sharp within the scale of Lorentz spaces. From this lack of optimality there follows useful information.

**Lemma 5.1.** *For any  $d \geq 2$  there exists  $C < \infty$  with the following property. Let  $\varepsilon > 0$ . Let  $f$  be any nonnegative measurable function in  $L^p(\mathbb{R}^d)$ . Then there exist an index set  $S \subset \mathbb{Z}$  of cardinality  $|S| \leq C\varepsilon^{-C}$  and a function  $\tilde{f}$  satisfying  $0 \leq \tilde{f} \leq f$  with rough level set decomposition  $\tilde{f} = \sum_{j \in S} 2^j f_j$ , such that*

$$(5.1) \quad \|\mathbf{T}\tilde{f}\|_q \geq (1 - \varepsilon)\|\mathbf{T}f\|_q.$$

*Proof.* Choose any  $r \in (p, q)$ . Let  $f$  have a rough level set decomposition  $f \equiv \sum_{j \in \mathbb{Z}} 2^j f_j$ ,  $f_j \leftrightarrow E_j$ . Let  $\eta > 0$  be a small parameter. Define

$$S = \{j : 2^j |E_j|^{1/p} > \eta\} \text{ and } \tilde{f} = \sum_{j \in S} 2^j f_j.$$

Then

$$\begin{aligned} \|f - \tilde{f}\|_{p,r}^r &= \sum_{j \notin S} (2^j |E_j|^{1/p})^r = \sum_{j \notin S} (2^j |E_j|^{1/p})^p (2^j |E_j|^{1/p})^{r-p} \\ &\leq \eta^{r-p} \sum_{j \in \mathbb{Z}} (2^j |E_j|^{1/p})^p = \eta^{r-p} \|f\|_p^p \end{aligned}$$

by Hölder's inequality. Therefore

$$\|\mathbf{T}(f - \tilde{f})\|_q \leq C \|f - \tilde{f}\|_{p,r} \leq C \eta^{1-p/r} \|f\|_p^{p/r}$$

where  $C < \infty$  is the norm of  $\mathbf{T}$  as an operator from  $L^{p,r}$  to  $L^q$ . Moreover,

$$\eta^p |S| = \sum_{j \in S} \eta^p \leq \sum_j 2^{jp} |E_j| \leq \|f\|_p^p.$$

We may assume without loss of generality that  $\|f\|_p = 1$ . Then defining  $\eta$  to satisfy  $C \eta^{1-p/r} = \varepsilon$  gives the conclusion stated.  $\square$

**Lemma 5.2.** *Let  $f \geq 0$  satisfy  $\|\mathbf{T}f\|_q \geq (1 - \delta)\mathbf{A}\|f\|_p$ . Then the function  $\tilde{f}$  in Lemma 5.1 can be chosen to satisfy*

$$(5.2) \quad \|f - \tilde{f}\|_p \leq C(\varepsilon + \delta)^{1/p} \|f\|_p$$

*in addition to all the conclusions of Lemma 5.1.*

*Proof.* Construct  $\tilde{f}$  as in the proof of Lemma 5.1. Since

$$\|\mathbf{T}\tilde{f}\|_q \leq \mathbf{A}\|\tilde{f}\|_p$$

and

$$\|\mathbf{T}\tilde{f}\|_q \geq (1 - \varepsilon)\|\mathbf{T}f\|_q,$$

it follows that

$$\|\tilde{f}\|_p \geq \mathbf{A}^{-1}(1 - \varepsilon)\|\mathbf{T}f\|_q.$$

Since  $\tilde{f}, f - \tilde{f}$  have disjoint supports,

$$\begin{aligned} \|f - \tilde{f}\|_p^p &= \|f\|_p^p - \|\tilde{f}\|_p^p \\ &\leq \|f\|_p^p - (1 - \varepsilon)^p \mathbf{A}^{-p} \|\mathbf{T}f\|_q^p \\ &\leq \|f\|_p^p - (1 - \varepsilon)^p \mathbf{A}^{-p} (1 - \delta)^p \mathbf{A}^p \|f\|_p^p \\ &= [1 - (1 - \varepsilon)^p (1 - \delta)^p] \|f\|_p^p \\ &\leq (C\varepsilon + C\delta) \|f\|_p^p. \end{aligned}$$

$\square$

The upshot is that near-extremals have low entropy, in the sense that relatively few terms in their rough level set decompositions suffice to approximate them to a specified degree of accuracy.

We have implicitly also established the following variant of Lemma 5.2.

**Lemma 5.3.** *There exists  $c, C \in \mathbb{R}^+$  with the following property. Suppose that  $0 \leq f \in L^p$ , and let  $f$  satisfy  $\|\mathbf{T}f\|_q \geq (1 - \delta)\mathbf{A}\|f\|_p$ , and have rough level set decomposition  $f = \sum_j 2^j f_j$ ,  $f_j \leftrightarrow E_j$ . Then for any  $\eta \in (0, 1]$ ,*

$$\left\| \sum_{j: 2^j |E_j|^{1/p} < \eta \|f\|_p} 2^j f_j \right\|_p \leq C(\delta^{1/p} + \eta^c) \|f\|_p.$$

## 6. STEP 2: WEAK HIGHER INTEGRABILITY

In Lemma 5.1 we obtained a very weak form of precompactness, in the form of an *a priori* bound on the cardinality of the index set  $S$  in the sum  $\tilde{f} = \sum_{j \in S} 2^j f_j$ . The next step is to show that the indices in  $S$  cannot be far apart from one another.

**Lemma 6.1.** *There exist constants  $c, C, \tilde{C} \in (0, \infty)$ , depending only on the dimension  $d$ , with the following property. Let  $\rho \in (0, 1)$ . Let  $f$  be a  $(1 - \delta)$ -quasiextremal for (1.2) satisfying  $\|f\|_p = 1$ . If  $\delta \leq c\rho^C$ , then there exists a function  $\tilde{f}$  satisfying  $\|f - \tilde{f}\|_p \leq C\rho^c$  with a rough level set decomposition  $\tilde{f} = \sum_j 2^j f_j$  such that if both  $\|2^i f_i\|_p \geq \rho$  and  $\|2^j f_j\|_p \geq \rho$ , then*

$$|i - j| \leq \tilde{C}\rho^{-\tilde{C}}$$

*Proof.* By Lemma 5.1 there exists a function  $\tilde{f}$  with rough level set decomposition  $\tilde{f} = \sum_{j \in S} 2^j f_j$ ,  $f_j \leftrightarrow E_j$ , with  $|S| \leq C\rho^{-C}$  such that  $\|f - \tilde{f}\|_p \leq C\rho^c$ , and moreover  $\|2^j f_j\|_p \geq \rho$  for every  $j \in S$ .

The operator dual to  $\mathbf{T}$  is identical to  $\mathbf{T}$  under conjugation by a simple change of variables. Thus Lemma 5.1 applies equally well to it. Therefore by duality, there exists  $h \in L^p$  satisfying  $\|h\|_p = 1$ , with rough level set decomposition  $h = \sum_{k \in \tilde{S}} 2^k h_k$ ,  $h_k \leftrightarrow F_k$ , with index set  $\tilde{S} \subset \mathbb{Z}$  satisfying  $|\tilde{S}| \leq C\rho^{-C}$ , so that

$$\langle h, \mathbf{T}\tilde{f} \rangle \geq (1 - 2\delta)\mathbf{A}.$$

Set  $N = |S| + |\tilde{S}| \leq C\rho^{-C}$ .

Set

$$M = \max_{i, j \in S} |i - j|.$$

It is possible to partition  $S$  into two nonempty disjoint sets  $S = S^\sharp \cup S^\flat$  so that  $|i - j| \geq M/N$  whenever  $i \in S^\sharp$  and  $j \in S^\flat$ . Fix any such partition.

Let  $\eta > 0$  be a small parameter to be chosen below; it will depend on  $N$  and thereby ultimately on  $\rho$ . Partition each of the sets  $F_k$  as  $F_k = F_k^\sharp \cup F_k^\flat \cup F_k^\natural$  measurably, so that

- (i) For each  $x \in F_k^\sharp$  there exists  $j \in S^\sharp$  such that  $\mathbf{T}\chi_{E_j}(x) > \eta|F_k|^{-1+1/p}|E_j|^{1/p}$ ;
- (ii) For each  $x \in F_k^\flat$  there exists  $j \in S^\flat$  such that  $\mathbf{T}\chi_{E_j}(x) > \eta|F_k|^{-1+1/p}|E_j|^{1/p}$ ;
- (iii) For each  $x \in F_k^\natural$ ,  $\mathbf{T}\chi_{E_j}(x) \leq \eta|F_k|^{-1+1/p}|E_j|^{1/p}$  for every  $j \in S$ .

Set

$$\begin{aligned} h^{\flat} &= \sum_{k \in \tilde{S}} 2^k h_k \chi_{F_k^{\flat}}, & h^{\sharp} &= \sum_{k \in \tilde{S}} 2^k h_k \chi_{F_k^{\sharp}}, \\ f^{\flat} &= \sum_{j \in S^{\flat}} 2^j f_j, & f^{\sharp} &= \sum_{j \in S^{\sharp}} 2^j f_j. \end{aligned}$$

For any  $j \in S$  and  $k \in \tilde{S}$ ,

$$\begin{aligned} \langle 2^k h_k \cdot \chi_{F_k^{\sharp}}, \mathbf{T}(2^j f_j) \rangle &\leq 2^{k+j+2} \langle \chi_{F_k^{\sharp}}, \mathbf{T} \chi_{E_j} \rangle \\ &\leq 2^{k+j+2} \eta |F_k^{\sharp}| \cdot |F_k|^{-1+1/p} |E_j|^{1/p} \\ &\leq 2^{k+j+2} \eta |F_k|^{1/p} |E_j|^{1/p} \\ &\leq 4\eta \|h\|_p \|f\|_p \\ &= 4\eta. \end{aligned}$$

$\tilde{f} = f^{\sharp} + f^{\flat}$  where  $f^{\sharp}, f^{\flat}$  have disjoint supports; likewise  $h = h^{\sharp} + h^{\flat} + h^{\natural}$ , with disjointly supported summands. Consequently, as in the proof of Lemma 5.1,

$$\langle h^{\sharp}, \mathbf{T} f^{\sharp} \rangle + \langle h^{\flat}, \mathbf{T} f^{\flat} \rangle \leq \mathbf{A}(1 - C\rho^c) \|f\|_p \|h\|_p = \mathbf{A}(1 - C\rho^c)$$

since  $\|f^{\sharp}\|_p \geq \rho$ ,  $\|f^{\flat}\|_p \geq \rho$ ,  $1 = \|h\|_p^p \geq \|h^{\sharp}\|_p^p + \|h^{\flat}\|_p^p$ , and  $1 = \|f\|_p^p \geq \|\tilde{f}\|_p^p = \|f^{\sharp}\|_p^p + \|f^{\flat}\|_p^p$ . Thus

$$\begin{aligned} \langle h, \mathbf{T} \tilde{f} \rangle &\leq CN^2 \eta + \langle h^{\sharp}, \mathbf{T} f^{\sharp} \rangle + \langle h^{\flat}, \mathbf{T} f^{\flat} \rangle + \langle h^{\natural}, \mathbf{T} f^{\sharp} \rangle + \langle h^{\natural}, \mathbf{T} f^{\flat} \rangle \\ &\leq CN^2 \eta + \mathbf{A}(1 - C\rho^c) + \langle h^{\flat}, \mathbf{T} f^{\sharp} \rangle + \langle h^{\sharp}, \mathbf{T} f^{\flat} \rangle. \end{aligned}$$

Set

$$(6.1) \quad \eta = c_0 \rho^{C_0},$$

where  $c_0, C_0$  are respectively sufficiently small and sufficiently large constants. In particular, choose  $c_0, C_0$  so that  $CN^2 \eta + (\mathbf{A} - C\rho^c) < (1 - 3\delta)\mathbf{A}$ ; this is possible since  $N \leq C\rho^{-C}$ . Thus

$$(6.2) \quad \langle h^{\flat}, \mathbf{T} f^{\sharp} \rangle \geq c\rho^C,$$

or

$$\langle h^{\sharp}, \mathbf{T} f^{\flat} \rangle \geq c\rho^C.$$

It is no loss of generality to assume (6.2), since the situation is symmetric with respect to interchange of the indices  $\sharp, \flat$ .

There must exist  $k \in \tilde{S}$  and  $j \in S^{\sharp}$  such that

$$(6.3) \quad \langle \chi_{F_k^{\flat}}, \mathbf{T} \chi_{E_j} \rangle \geq \eta |F_k|^{1/p} |E_j|^{1/p};$$

otherwise we would have a total bound of  $C\eta N^2$  for the left-hand side in (6.2), which would be a contradiction due to the choice of  $\eta$ .

Let

$$F_k^{\flat, \dagger} = \{x \in F_k^{\flat} : T\chi_{E_j}(x) < \tfrac{1}{2}\eta |F_k^{\flat}|^{-1} |F_k|^{1/p} |E_j|^{1/p}\}.$$



Then

$$\langle \chi_{F_k^b, \dagger}, T\chi_{E_j} \rangle \leq \frac{1}{2}\eta|F_k|^{1/p}|E_j|^{1/p},$$

so

$$\langle \chi_{F_k^b \setminus F_k^{b, \dagger}}, T\chi_{E_j} \rangle \geq \frac{1}{2}\eta|F_k|^{1/p}|E_j|^{1/p}.$$

Since

$$\langle \chi_{F_k^b \setminus F_k^{b, \dagger}}, T\chi_{E_j} \rangle \leq \mathbf{A}|F_k^b \setminus F_k^{b, \dagger}|^{1/p}|E_j|^{1/p},$$

we must have

$$|F_k^b \setminus F_k^{b, \dagger}| \geq c\rho^C|F_k|.$$

Thus

$$T\chi_{E_j}(x) \geq c\rho^C|F_k|^{-1+1/p}|E_j|^{1/p}$$

for every  $x \in F_k^b \setminus F_k^{b, \dagger}$ .

There exist an index  $i \in S^b$  and a subset  $\mathcal{F} \subset F_k^b \setminus F_k^{b, \dagger}$  such that  $\mathbf{T}\chi_{E_i}(x) > \eta|F_k|^{-1+1/p}|E_i|^{1/p}$  for all  $x \in \mathcal{F}$  and

$$\langle \chi_{\mathcal{F}}, \mathbf{T}\chi_{E_j} \rangle \geq N^{-1}\langle \chi_{F_k^b \setminus F_k^{b, \dagger}}, \mathbf{T}\chi_{E_j} \rangle \geq cN^{-1}\rho^C|F_k|^{1/p}|E_j|^{1/p}.$$

Since  $\langle \chi_{\mathcal{F}}, \mathbf{T}\chi_{E_j} \rangle \leq \mathbf{A}|\mathcal{F}|^{1/p}|E_j|^{1/p}$ , it follows again that  $|\mathcal{F}| \geq c\rho^C|F_k|$ .

Therefore

$$\begin{aligned} T\chi_{E_j}(x) &\geq c\rho^C|\mathcal{F}|^{-1+1/p}|E_j|^{1/p} \\ T\chi_{E_i}(x) &\geq c\rho^C|\mathcal{F}|^{-1+1/p}|E_i|^{1/p} \end{aligned}$$

for every  $x \in \mathcal{F}$ . By Lemma 3.13, this forces

$$|E_i| \leq C\rho^{-C}|E_j| \text{ and } |E_j| \leq C\rho^{-C}|E_i|.$$

Since  $\|f_i\|_p \geq \rho$ ,  $\rho^p \leq 2^p 2^{pi}|E_i|$ . On the other hand,  $2^{p(j+1)}|E_j| \leq \|\tilde{f}\|_p^p = 1$ . Therefore

$$|E_j| \leq C2^{(i-j)p}\rho^{-p}|E_i|.$$

In conjunction with the reverse bound  $|E_j| \geq c\rho^C|E_i|$  proved above, this forces  $|i - j| \leq C \log(1/\rho)$ .

This conclusion holds for a certain pair  $(i, j) \in S^b \times S^\sharp$ . The partition  $S = S^\sharp \cup S^b$  was chosen so that  $M \leq N|i - j|$ . Therefore  $M \leq C\rho^{-C}$ .  $\square$

**Remark 6.2.** This proof is on track to establish higher integrability in the natural form  $\phi^*f \in L^Q$  for some  $Q > p$  for extremals, until the very last step, in which  $|i - j|$  turns into  $N|i - j|$ . Perhaps some more efficient reorganization is possible.

**Corollary 6.3.** *There exist a finite constant  $C$  and a function  $\Psi : (0, \infty) \rightarrow (0, \infty)$  satisfying  $\Psi(t)/t^p \rightarrow \infty$  as  $t \rightarrow \infty$  and as  $t \rightarrow 0$ , with the following property. For any  $\varepsilon > 0$  there exists  $\delta > 0$  such that for any nonnegative function  $f$  satisfying  $\|f\|_p = 1$  and  $\|\mathbf{T}f\|_q \geq (1 - \delta)\mathbf{A}$ , there exist  $\phi \in \mathcal{G}_d$  and a decomposition  $\phi^*f = g + h$  such that  $g, h \geq 0$  satisfy  $\|h\|_p < \varepsilon$  and*

$$(6.4) \quad \int \Psi(g) \leq C.$$

*Proof.* Fix any exponent  $r \in (p, q)$ . Let  $\eta > 0$  be a small quantity to be chosen at the end of the proof. Let  $\delta$  be a sufficiently small function of  $\eta, \varepsilon$ .

Let  $f$  have a rough level set decomposition  $f \equiv \sum_{j \in \mathbb{Z}} 2^j f_j$ ,  $f_j \leftrightarrow E_j$ . By Lemma 5.3, if  $\delta > 0$  is sufficiently small then there exists at least one index  $k$  such that  $\|2^k f_k\|_p \geq c_0$ , where  $c_0 > 0$  is an absolute constant which depends only on the dimension  $d$ . By choosing  $\phi \in \mathcal{G}_d$  to be an appropriate dilation symmetry  $\phi(x', x_d) = (rx', r^2 x_d)$  we may reduce to the case where  $k = 0$ .

By Lemma 6.1, there exists  $M < \infty$  such that  $\|2^j f_j\|_p < \eta$  whenever  $|j| \geq M$ . Define  $h = \sum_{|j| > M} 2^j f_j$ , and  $g = \sum_{|j| \leq M} 2^j f_j$ . Then  $\|h\|_{L^{p,r}}^r = \sum_{|j| > M} \|2^j f_j\|_p^r \leq \eta^{r-p} \sum_{|j| > M} \|2^j f_j\|_p^p \leq \eta^{r-p}$ . By the proof of Lemma 5.2, this implies that  $\|h\|_p < \varepsilon$  provided that  $\delta, \eta$  are chosen to be sufficiently small.

If  $\eta, \delta$  are chosen to be sufficiently small then by Lemma 6.1, for any  $\rho \geq \eta$ ,  $|j| \leq C\rho^{-C}$  whenever  $\|2^j f_j\|_p \geq \rho$ . Since the sets  $E_j$  are pairwise disjoint, for any nondecreasing function  $\Psi$ ,  $\int \Psi(g) = \sum_{|j| \leq M} \int \Psi(2^j f_j) \leq \sum_{|j| \leq M} \Psi(2^{j+1}) |E_j|$ .

Let  $S_k = \{j : |j| \leq M \text{ and } \|2^j f_j\|_p \in (2^{-k-1}, 2^{-k}]\}$ . Then for all  $j \in S_k$ ,  $|j| \leq C2^{Ck}$  by Lemma 6.1. Therefore

$$\sum_{j \in S_k} \Psi(2^{j+1}) |E_j| \leq 2^p \max_{|j| \leq C2^{Ck}} \frac{\Psi(2^{j+1})}{2^{p(j+1)}} \cdot \sum_{j \in S_k} \|2^j f_j\|_p^p.$$

Moreover,

$$\sum_{j \in S_k} \|2^j f_j\|_p^p \leq C\delta + C2^{-ck}$$

by Lemma 5.3. Thus

$$\sum_{2^k \leq \eta^{-1}} \sum_{j \in S_k} \Psi(2^{j+1}) |E_j| \leq \sum_{2^k \leq \eta^{-1}} C(\delta + 2^{-ck}) \max_{|j| \leq C2^{Ck}} \left( \frac{\Psi(2^{j+1})}{2^{p(j+1)}} \right).$$

Choose  $\Psi$  to be a nondecreasing function satisfying the growth condition

$$\sum_{k=0}^{\infty} 2^{-ck} \max_{|j| \leq C2^{Ck}} \frac{\Psi(2^{j+1})}{2^{p(j+1)}} < \infty,$$

and  $\Psi(t)/t^p \rightarrow \infty$  as  $t \rightarrow 0$  and as  $t \rightarrow \infty$ . Then choose  $\delta$  to be a function of  $\eta$ , satisfying

$$\delta \sum_{2^k \leq \eta^{-1}} \max_{|j| \leq C2^{Ck}} \frac{\Psi(2^{j+1})}{2^{p(j+1)}} \leq 1.$$

This completes the proof.  $\square$

## 7. STEP 3: SPATIAL LOCALIZATION

Lemma 3.10 leads to the following preliminary result concerning the geometric structure of quasiextremals.

**Lemma 7.1.** *For any  $\varepsilon > 0$  there exist  $\delta > 0$  and  $N, K < \infty$  with the following property. Let  $f \geq 0$  be any  $(1 - \delta)$ -quasiextremal for inequality (1.2) satisfying*

$\|f\|_p = 1$ . Then there exists a function  $F$  with a rough level set decomposition  $F = \sum_{j \in S} 2^j F_j$ ,  $F_j \leftrightarrow E_j$ , satisfying

$$\begin{aligned} 0 &\leq F \leq f \\ \|\mathbf{T}F\|_q &\geq (1 - \varepsilon)\mathbf{A}, \\ |i - j| &\leq K \text{ for all } i, j \in S, \end{aligned}$$

and for each  $j \in S$  there exist  $N$  paraballs  $B_{j,i}$  such that

$$(7.1) \quad E_j \subset \cup_{i=1}^N B_{j,i}$$

$$(7.2) \quad \sum_i |B_{j,i}| \leq C(\varepsilon)|E_j|.$$

*Proof.* In light of results already proved, by modifying  $f$  by a function whose  $L^p$  norm is  $< \varepsilon/2$ , we may suppose that  $f$  has a finite rough level set decomposition  $f = \sum_{j \in S} 2^j f_j$ ,  $f_j \leftrightarrow \tilde{E}_j$ , with  $|i - j| \leq K(\varepsilon)$  for all  $i, j \in S$ . Let  $\eta > 0$  be a small quantity. Then  $\|\mathbf{T}f\|_q \geq (1 - \delta)\mathbf{A} \geq \eta$ , so we may apply Lemma 3.10 to find a paraball  $B$  and an index  $i_1$  such that  $\|2^{i_1} f_{i_1} \chi_B\|_p \geq \rho$  and  $|B| \leq |\tilde{E}_{i_1}|$ , where  $\rho > 0$  depends only on  $\eta$ . Set  $g_1 = 2^{i_1} f_{i_1} \chi_B$  and write  $f = g_1 + h_1$ .

This was step 1 of a construction which we iterate, as follows. At step  $n$ , one is given a collection of paraballs  $\{B_m : m \leq n-1\}$ , along with a decomposition

$$f = \sum_{m=1}^{n-1} g_m + h_{n-1}$$

such that the sets  $\tilde{B}_m = B_m \setminus \cup_{l < m} B_l$  and functions  $g_m, h_{n-1}$  satisfy

$$g_m = 2^{i_m} f_{i_m} \chi_{\tilde{B}_m} \text{ for some } i_m \in \mathbb{Z},$$

$$|B_m| \leq |\tilde{E}_{i_m}|,$$

$$h_{n-1} = f \cdot \chi_{\mathbb{R}^d \setminus \cup_{m \leq n-1} B_m}.$$

Since  $0 \leq h_{n-1} \leq f$ ,  $\|h_{n-1}\|_p \leq \|f\|_p \leq 1$ . If  $\|\mathbf{T}h_{n-1}\|_q < \eta$  then the construction terminates. Otherwise invoke Lemma 3.10 to find a paraball  $B_n$  and an index  $i_n$  such that the function

$$g_n = 2^{i_n} f_{i_n} \cdot \chi_{\mathbb{R}^d \setminus \cup_{m < n} B_m} \cdot \chi_{B_n}$$

satisfies  $\|g_n\|_p \geq \rho$  and  $|B_n| \leq |\tilde{E}_{i_n}|$ .

The functions  $g_n$  are nonnegative and have pairwise disjoint supports, and  $\sum_{m=1}^n g_m \leq f$ . Consequently  $\sum_{m=1}^n \|g_m\|_p^p \leq \|f\|_p^p = 1$ , so this process must terminate after at most  $\rho^{-p}$  iterations. If it terminates at the  $n$ -th step, then set  $F = \sum_{m=1}^{n-1} g_m$ . Then  $\|\mathbf{T}(f - F)\|_q \leq \eta$ , so by Lemma 5.2,  $\|f - F\|_p < \frac{1}{2}\varepsilon$  provided that  $\eta$  is chosen to be sufficiently small. Defining the collection  $\{B_{j,i}\}$  to be  $\{\tilde{B}_m : i_m = j\}$  for each index  $j$  produces a collection of paraballs satisfying (7.1), (7.2).  $\square$

Consider momentarily the possibility of a sequence  $\{f_\nu\}$  of quasiextremals which are characteristic functions of sets  $E_\nu$  satisfying  $|E_\nu| = 1$ . If these sets were to move off to spatial infinity as  $\nu \rightarrow \infty$ , then  $f_\nu$  and  $f_\nu^p$  would converge weakly to zero,

preventing the extraction of an extremal as a limit of some subsequence. If each  $E_\nu$  were a parabol, then this situation could be rectified by invoking the symmetry group  $\mathcal{G}_d$  to replace each  $E_\nu$  by a parabol independent of  $\nu$ . However, there is potentially a more problematic obstruction: If each  $E_\nu$  were a disjoint union  $E_\nu = E'_\nu \cup E''_\nu$ , with  $|E'_\nu| = |E''_\nu| = \frac{1}{2}$  and with  $E'_\nu, E''_\nu$  moving to infinity in different directions, then the symmetries would not suffice to produce a useful renormalized sequence. The following refinement of Lemma 7.1 rules out this sort of obstruction.

**Lemma 7.2.** *For any  $\varepsilon > 0$  there exist  $\delta > 0$  and  $K, \lambda < \infty$  with the following properties. Let  $f \geq 0$  be any  $(1 - \delta)$ -quasiextremal for inequality (1.2) satisfying  $\|f\|_p = 1$ . Then there exist a function  $\tilde{f}$  and a parabol  $B$  such that*

$$\begin{aligned} 0 &\leq \tilde{f} \leq f, \\ \|\tilde{f}\|_p &\geq 1 - \varepsilon, \\ \|T\tilde{f}\|_q &\geq (1 - \varepsilon)\mathbf{A}, \end{aligned}$$

and  $\tilde{f}$  admits a rough level set decomposition  $\tilde{f} = \sum_{j \in S} 2^j f_j$  with a distinguished index  $J \in \mathbb{Z}$  satisfying

$$\begin{aligned} |j - J| &\leq K \text{ for all } j \in S, \\ f_j &\text{ is supported in } \lambda B \text{ for all } j \in S, \\ 2^J |B|^{1/p} &\leq C \|f\|_p. \end{aligned}$$

This improves upon Lemma 7.1 in that the collection of paraballs  $B_{j,i}$  in the conclusion has been replaced by a single parabol, which however must be expanded by the factor  $\lambda$ . This is a geometric analogue of the replacement of an upper bound on the cardinality of the index set  $S$  in Lemma 5.2 by an upper bound on its diameter in Lemma 6.1. Lemma 7.2 follows directly from the combination of Lemmas 4.1 and 7.1, in the same way that Lemma 6.1 followed from Lemma 5.2 combined with Lemma 3.13. Details are left to the reader.

## 8. WEAK CONVERGENCE AND EXTREMIZERS

Let  $f$  be a  $(1 - \delta)$ -quasiextremal satisfying  $\|f\|_p = 1$  for  $\delta$  sufficiently small. Apply Lemma 7.2. Then by replacing  $f$  by  $\phi^* f$  for an appropriately chosen  $\phi \in \mathcal{G}_d$ , we may reduce to the case where in the conclusions of that lemma,  $J = 0$  and the parabol  $B$  is  $B = \{x \in \mathbb{R}^d : |x_j| \leq 1 \text{ for all } 1 \leq j \leq d - 1 \text{ and } |x_d - |x'|^2| < 1\}$ . Thus we have the following information.

**Lemma 8.1.** *There exist a constant  $C_0 < \infty$  and functions  $\Psi : [0, \infty) \rightarrow [0, \infty)$  and  $\rho : [1, \infty) \rightarrow (0, \infty)$  satisfying  $\Psi(t) \geq t^p$  for all  $t$ ,  $\Psi(t)/t^p \rightarrow \infty$  as  $t \rightarrow 0$  and as  $t \rightarrow \infty$ , and  $\rho(R) \rightarrow 0$  as  $R \rightarrow \infty$ , such that for any  $\varepsilon > 0$  there exists  $\delta > 0$  with the following property. Let  $f$  be any nonnegative function satisfying  $\|f\|_p = 1$  and  $\|\mathbf{T}f\|_q \geq (1 - \delta)\mathbf{A}$ . Then there exist  $\phi \in \mathcal{G}_d$  and a decomposition*

$$\phi^* f = g + h$$

with  $g, h \geq 0$  satisfying

$$\|h\|_p < \varepsilon, \quad \int_{\mathbb{R}^d} \Psi(g) \leq C_0, \quad \int_{|x| \geq R} g^p \leq \rho(R).$$

Moreover, there exists a nonnegative function  $F$  satisfying  $\|F\|_p = 1$  and

$$\langle F, \mathbf{T}g \rangle \geq (1 - \varepsilon)\mathbf{A}, \quad \int_{\mathbb{R}^d} \Psi(F) \leq C_0, \quad \int_{|x| \geq R} F^p \leq \rho(R).$$

The bound for  $\int \Psi(g)$  is Corollary 6.3. The bound for  $\int_{|x| \geq R} g^p$  follows in a similar way from Lemma 7.2.

With these uniform bounds in hand, it is straightforward to derive most of the conclusions of Theorem 2.2. For any sequence of functions  $\{f_\nu\}$ , we write  $f_\nu \rightharpoonup f$  to mean that  $\int f_\nu \varphi \rightarrow \int f \varphi$  as  $\nu \rightarrow \infty$ , for every compactly supported continuous test functions  $\varphi$ . If  $f_\nu$  are nonnegative  $L^1$  functions, then  $f_\nu \rightharpoonup f$  implies  $\|f_\nu\|_1 \rightarrow \|f\|_1$ , provided that  $\sup_\nu \int_{|x| \geq R} f_\nu \rightarrow 0$  as  $R \rightarrow \infty$ ,

*Proof of existence of extremizers.* Let  $\{f_\nu\}$  be any extremizing sequence. Then by the preceding lemma, there exist  $\phi_\nu \in \mathcal{G}_d$  such that  $\phi_\nu^* f_\nu = g_\nu + h_\nu$  where  $\|h_\nu\|_p \rightarrow 0$ , while the functions  $g_\nu$  satisfy the other conclusions of Lemma 8.1; and there exist corresponding functions  $F_\nu$  satisfying the same bounds as  $g_\nu$ , with  $\|F_\nu\|_p = 1$  and  $\langle F_\nu, \mathbf{T}g_\nu \rangle \rightarrow \mathbf{A}$ .

It follows directly from the Banach-Alaoglu theorem that after passage to a subsequence of the index  $\nu$ ,  $F_\nu^p \rightharpoonup F^p$  and  $g_\nu^p \rightharpoonup g^p$  form some  $F, g \in L^p$ . Since  $\int F_\nu^p \eta \leq 1$  for every continuous, compactly supported function  $\eta$  satisfying  $\|\eta\|_\infty \leq 1$ , necessarily  $\int F^p \leq 1$ ; likewise  $\int g^p \leq 1$ .

We claim that

$$(8.1) \quad \langle F_\nu, \mathbf{T}g_\nu \rangle \rightarrow \langle F, \mathbf{T}g \rangle.$$

Since  $\|F\|_p \leq 1$  and likewise  $\|g\|_p \leq 1$  and  $\langle F_\nu, \mathbf{T}g_\nu \rangle \rightarrow \mathbf{A}$ , it follows that  $\|F\|_p = \|g\|_p = 1$  and that  $g$  is an extremal.

To prove (8.1) define

$$g_{\nu,\lambda}(x) = g_\nu(x) \chi_{|x| \leq \lambda}(x) \chi_{g_\nu(x) \leq \lambda}(x),$$

and define  $g^{(\lambda)}, F_{\nu,\lambda}, F^{(\lambda)}$  in the corresponding way. For any compactly supported function  $\eta \in C_0^1(\mathbb{R}^d)$ , the operator  $f \mapsto \eta \mathbf{T} \eta f$  is smoothing in the sense that it maps  $L^2(\mathbb{R}^d)$  boundedly to the Sobolev space  $H^s(\mathbb{R}^d)$  for  $s = (d-1)/2$ .  $H^s$  embeds compactly into  $L^2$  in any bounded region. Therefore the weak  $L^p$  convergence of  $g_{\nu,\lambda}$  to  $g^{(\lambda)}$  as  $\nu \rightarrow \infty$  implies  $L^2$  norm convergence of  $\mathbf{T}(g_{\nu,\lambda})$  to  $\mathbf{T}(g^{(\lambda)})$  as  $\nu \rightarrow \infty$ , for every fixed  $\lambda$ . Therefore  $\langle F_{\nu,\lambda}, \mathbf{T}g_{\nu,\lambda} \rangle \rightarrow \langle F^{(\lambda)}, \mathbf{T}g^{(\lambda)} \rangle$  as  $\nu \rightarrow \infty$ , for every fixed  $\lambda < \infty$ .

The conclusions of Lemma 8.1 guarantee that  $g_{\nu,\lambda} \rightarrow g_\nu$  and  $F_{\nu,\lambda} \rightarrow F_\nu$  in  $L^p$  norm uniformly in  $\nu$  as  $\lambda \rightarrow \infty$ . This uniform convergence, together with the convergence proved in the preceding paragraph, give (8.1).  $\square$

$L^p$  norm convergence of  $g_\nu$  to  $g$  will be proved below.

*Proof of part (iv) of Theorem 2.2.* Let  $f$  be an arbitrary nonnegative extremizer satisfying  $\|f\|_p = 1$ . For any  $n$  there exist  $\phi_n \in \mathcal{G}_d$  and a decomposition  $\phi_n^* f = g_n + h_n$  where  $g_n$  satisfies the inequalities of Lemma 8.1, and  $\|h_n\|_p < 2^{-n}$ . By passing to a subsequence, we may arrange that  $g_n$  converges in  $L^p$  and almost everywhere to a limit  $g$ , which is an extremizer. Since  $\phi_n^*$  preserves the  $L^p$  norm and  $\|h_n\|_p \rightarrow 0$ ,  $\phi_n^* f \rightarrow g$  and  $(\phi_n^{-1})^* g \rightarrow f$  in  $L^p$  norm. It is an elementary consequence of the concrete description of  $\mathcal{G}_d$  that this forces  $\{\phi_n^{-1}\}$  to be a precompact family of diffeomorphisms of  $\mathbb{R}^d$ . By passing to a subsequence we conclude that  $f = \phi^* g$  for some  $\phi \in \mathcal{G}_d$ . Thus  $f$  satisfies all of the conclusions stated in conclusion (iv) of Theorem 2.2.  $\square$

## 9. LOOSE ENDS

*Proof of the Euler-Lagrange identity (2.7).* Let  $f$  be any nonnegative extremizer satisfying  $\|f\|_p = 1$ . Since  $q = d + 1$  is the exponent conjugate to  $p = (d + 1)/d$ ,  $\mathbf{T}^*$  has the same norm as  $\mathbf{T}$ , as operators from  $L^p$  to  $L^q$ . Then

$$\begin{aligned} \mathbf{A}^{d+1} \|f\|_p^{d+1} &= \langle \mathbf{T}f, (\mathbf{T}f)^d \rangle = \langle f, \mathbf{T}^*([\mathbf{T}f]^d) \rangle \\ &\leq \|f\|_p \|T^*(\mathbf{T}f)^d\|_q \leq \mathbf{A} \|f\|_p \|(\mathbf{T}f)^d\|_p \\ &= \mathbf{A} \|f\|_p \|\mathbf{T}f\|_q^d = \mathbf{A}^{d+1} \|f\|_p^{d+1}. \end{aligned}$$

The first inequality is an application of Hölder's inequality. Since the extreme left- and right-hand sides are equal, both inequalities in this chain must be equalities. Equality in Hölder's inequality forces  $\mathbf{T}^*([\mathbf{T}f]^d)$  to agree almost everywhere with some constant multiple of  $f^{1/d}$ . The chain of equalities then forces this constant to be  $\mathbf{A}^{d+1}$ .  $\square$

**Lemma 9.1.** *If  $f$  is a nonnegative extremizer for (1.2), then  $f(x) > 0$  for almost every  $x \in \mathbb{R}^d$ .*

*Proof.* Let  $f$  be a nonnegative extremizer satisfying  $\|f\|_p = 1$ . Consider initially the set  $G$  consisting of all points expressible in the form  $z + (s', |s'|^2) - (t', |t'|^2)$  such that  $z$  is a Lebesgue point of  $\{y : f(y) > 0\}$  and  $s' \neq t' \in \mathbb{R}^{d-1}$ . We claim that  $f(x) > 0$  for almost every  $x \in G$ . If not, choose some subset  $\tilde{G} \subset G$  satisfying  $0 < |\tilde{G}| < \infty$ . Consider the functions  $f_\varepsilon = f + \varepsilon \chi_{\tilde{G}}$ . Then  $\|f_\varepsilon\|_p^p = 1 + O(\varepsilon^p) = 1 + o(\varepsilon)$  as  $\varepsilon \rightarrow 0$ . On the other hand,

$$\|\mathbf{T}f_\varepsilon\|_{d+1}^{d+1} \geq \|\mathbf{T}f\|_{d+1}^{d+1} + (d+1)\varepsilon \int (\mathbf{T}f)^d \mathbf{T}\chi_{\tilde{G}}.$$

Now because  $f, \chi_{\tilde{G}}$  are nonnegative functions and  $\mathbf{T}$  is defined by convolution with a nonnegative measure,  $\int (\mathbf{T}f)^d \mathbf{T}\chi_{\tilde{G}} > 0$  if and only if  $\int \mathbf{T}f \cdot \mathbf{T}\chi_{\tilde{G}} = \langle \mathbf{T}^* \mathbf{T}f, \chi_{\tilde{G}} \rangle > 0$ .

$\mathbf{T}^* \circ \mathbf{T}$  is defined by convolution with a measure which has a continuous, strictly positive Radon-Nikodym derivative with respect to Lebesgue measure, on the open set  $\{(s', |s'|^2) - (t', |t'|^2) : s' \neq t'\}$ . The condition that  $\tilde{G} \subset G$  thus ensures that  $\mathbf{T}^* \mathbf{T}f > 0$  at every point of  $\tilde{G}$ . Therefore  $\langle \mathbf{T}^* \mathbf{T}f, \chi_{\tilde{G}} \rangle > 0$ .

Consequently  $\|\mathbf{T}f_\varepsilon\|_{d+1} \geq \mathbf{A} + c\varepsilon$  for some  $c > 0$  for all sufficiently small  $\varepsilon > 0$ , and therefore  $\|\mathbf{T}f_\varepsilon\|_q / \|f_\varepsilon\|_p \geq \mathbf{A} + c'\varepsilon$  for small positive  $\varepsilon$  for all  $c' < c$ , contradicting

the extremality of  $f$ . Thus any extremizer must be positive almost everywhere on some open set.

This additional information can be fed back into the above argument, which then demonstrates that  $f > 0$  almost everywhere at every point  $z + (s', |s'|^2) - (t', |t'|^2)$  where  $z$  varies over a subset of full measure of some ball. One more iteration establishes the conclusion.  $\square$

**Corollary 9.2.** *Let  $f$  be a nonnegative extremizer for (1.2). Then for any compact set  $K \subset \mathbb{R}^d$  there exists  $c > 0$  such that  $f(x) \geq c$  for almost every  $x \in K$ .*

This follows from the Euler-Lagrange equation by reasoning already used in the proof of Lemma 9.1, since  $\mathbf{T}^* \circ \mathbf{T}$  is expressed by convolution with a nonnegative measure  $\mu$  with the property that any point is expressible as a finite sum of elements of an open set on which  $\mu$  has a continuous, strictly positive Radon-Nikodym derivative.

It remains to be proved that any extremizing sequence has a subsequence which converges in  $L^p$  norm, rather than the weak convergence proved above.

**Lemma 9.3.** *Let  $\{f_\nu\}$  be an extremizing sequence for inequality (1.2); thus  $f_\nu \geq 0$ ,  $\|f_\nu\|_p = 1$ , and  $\|\mathbf{T}f_\nu\|_q \rightarrow \mathbf{A}$ . Then there exist a sequence of symmetries  $\{\phi_\nu\} \subset \mathcal{G}_d$ , an extremal  $F$  for (1.2), and a subsequence  $\{f_{\nu_k}\}$  such that  $\phi_{\nu_k}^* f_{\nu_k} \rightarrow F$  in  $L^{(d+1)/d}(\mathbb{R}^d)$  norm.*

*Proof.* After passing to a subsequence, we may choose  $\phi_\nu$  so that  $\phi_\nu^* f_\nu = F_\nu + h_\nu$ , so that  $\|h_\nu\|_p \rightarrow 0$ ,  $\|F_\nu\|_p \rightarrow 1$ ,  $F_\nu^p \rightarrow F^p$  for some nonnegative extremal  $F$  satisfying  $\|F\|_p = 1$ , and  $F_\nu, F$  satisfy the higher integrability and spatial decay bounds provided by Lemma 8.1. Moreover, there exists  $H \geq 0$  satisfying  $\|H\|_p = 1$  and  $\mathbf{T}^* H = \mathbf{A} F^{p-1}$  almost everywhere.

As in the proof of (8.1) it follows from the *a priori* bounds and Rellich's lemma that after passing to a subsequence,  $\langle F_\nu, \mathbf{T}^* H \rangle \rightarrow \langle F, \mathbf{T}^* H \rangle = \mathbf{A} \langle F, F^{p-1} \rangle = \mathbf{A}$ . Therefore  $\langle F_\nu, F^{p-1} \rangle \rightarrow 1 = \langle F, F^{p-1} \rangle$ .

For any  $\nu$  and any small  $\delta > 0$  denote  $E_{\delta, \nu} = \{x : F_\nu^p(x) \leq (1 - \delta)F^p(x)\}$ . Recall that  $q = d + 1$  is the exponent dual to  $p = (d + 1)/d$ . Then

$$\begin{aligned} \int F_\nu F^{p-1} &\leq \int p^{-1} F_\nu^p + q^{-1} (F^{p-1})^q = \int p^{-1} F_\nu^p + q^{-1} F^p \\ &\leq p^{-1} \int F^p + q^{-1} \int F^p - \delta p^{-1} \int_{E_{\delta, \nu}} F^p = 1 - \delta p^{-1} \int_{E_{\delta, \nu}} F^p. \end{aligned}$$

Therefore

$$\int_{E_{\delta, \nu}} F^p \rightarrow 0 \text{ as } \nu \rightarrow \infty$$

Because  $F > 0$  almost everywhere, this implies that for any compact subset  $K \subset \mathbb{R}^d$ ,  $|E_{\delta, \nu} \cap K| \rightarrow 0$  as  $\nu \rightarrow \infty$ . Together with the higher integrability and spatial decay bounds of Lemma 8.1, this implies that  $\int_{E_{\delta, \nu}} F_\nu^p \rightarrow 0$  also. We may choose  $\delta = \delta(\nu)$  to tend to zero as  $\nu \rightarrow \infty$ , yet still have

$$\int_{E_{\delta(\nu), \nu}} (F^p + F_\nu^p) \rightarrow 0.$$

Now

$$\int_{F_\nu^p(x) \geq (1-\delta)F^p(x)} (F^p - F_\nu^p) \leq \delta \int F^p \leq \delta.$$

Therefore by splitting the integral into the two regions  $F_\nu^p \leq (1-\delta(\nu))F^p$  and  $(1-\delta(\nu))F^p \leq F_\nu^p \leq F^p$  we conclude that

$$(9.1) \quad \int_{F_\nu \leq F} (F^p - F_\nu^p) \rightarrow 0 \text{ for each } \delta > 0.$$

Since  $\int F_\nu^p = \int F^p$ , (9.1) forces  $\int |F^p - F_\nu^p| \rightarrow 0$ .  $\square$

One final conclusion of Theorem 2.2 remains to be established, the existence of a decomposition  $g = g^\sharp + g^\flat$  where  $\|g^\sharp\|_{C^1} \leq R$  and  $\|g^\flat\|_p \leq \eta(R)$ , where  $\eta \rightarrow 0$  as  $R \rightarrow \infty$ . Equivalently, if  $\zeta \in C^\infty(\mathbb{R}^d)$  is compactly supported and satisfies  $\zeta(0) = 1$ , and if  $M_\rho$  is the Fourier multiplier operator with symbol  $1 - \zeta(\xi/\rho)$  then  $\|M_\rho g\|_p \leq \tilde{\eta}(\rho)$  where  $\tilde{\eta}(\rho) \rightarrow 0$  as  $\rho \rightarrow \infty$ . Here  $g$  is a  $(1-\delta)$ -quasiextremal which satisfies (2.4) and (2.5), and  $\delta$  is permitted to depend on  $R$  and/or  $\rho$  and can be chosen to be as small as may be desired.

*Proof of (2.6).* Fix  $F$  satisfying  $\|F\|_p = 1$  such that  $\langle g, \mathbf{T}^* F \rangle \geq (1-\delta)\mathbf{A}$ . Let  $\varepsilon > 0$ . By the results proved above, if  $\rho$  is sufficiently large then  $\mathbf{T}^* F$  may be decomposed as  $\mathbf{T}^* F = H^\sharp + H^\flat$  where  $\widehat{H^\sharp}(\xi) = 0$  for all  $|\xi| \geq \rho/2$  and  $\|H^\flat\|_q < \varepsilon$ .

Split  $g = g^\sharp + g^\flat$  where  $\widehat{g^\sharp}$  is supported where  $|\xi| < 2\rho$ ,  $\widehat{g^\flat}$  is supported where  $|\xi| > \rho$ , and  $\|g^\sharp\|_p + \|g^\flat\|_p \leq C < \infty$ ; this may be done with  $C$  a constant depending only on  $d, p$ . To do this, set  $g^\flat = M_\rho g$  where  $M_\rho$  has symbol  $1 - \zeta(\xi/\rho)$ , with the auxiliary function  $\zeta \in C_0^\infty$  chosen to satisfy  $\zeta(\xi) \equiv 1$  whenever  $|\xi| \leq 2$ . Then  $g^\flat$  will continue to satisfy (2.4) and (2.5), uniformly in  $\rho, g$  so long as  $\rho \geq 1$ .

We aim to prove that if  $\delta, \varepsilon$  are sufficiently small and  $\rho$  is sufficiently large, then  $\|g^\flat\|_p$  is less than any preassigned quantity.

Consider  $\tilde{g} = g^\sharp - g^\flat$ . Then

$$\begin{aligned} \langle \tilde{g}, \mathbf{T}^* F \rangle &= \langle g, \mathbf{T}^* F \rangle - 2\langle g^\flat, \mathbf{T}^* F \rangle \\ &= \langle g, \mathbf{T}^* F \rangle - 2\langle g^\flat, H^\flat \rangle \\ &\geq (1-\delta)\mathbf{A} - 2\|g^\flat\|_p \|H^\flat\|_q \\ &\geq (1-\delta)\mathbf{A} - 2C\varepsilon. \end{aligned}$$

Thus

$$\|\tilde{g}\|_p \geq (1-\delta) - 2C\varepsilon\mathbf{A}^{-1}.$$

The same reasoning gives

$$(9.2) \quad \|g^\sharp\|_p \geq (1-\delta) - C\varepsilon\mathbf{A}^{-1}.$$

Since  $1 < p < 2$ , there exists  $c_p < \infty$  such that for any  $s, t \in \mathbb{R}$ ,

$$\frac{1}{2}|s+t|^p + \frac{1}{2}|s-t|^p \geq |s|^p + c_p \min(|s|^{p-2}|t|^2, |t|^p).$$

Therefore

$$\frac{1}{2}|g(x)|^p + \frac{1}{2}|\tilde{g}(x)|^p \geq |g^\sharp(x)|^p + c_p \min(|g^\sharp(x)|^{p-2}|g^\flat(x)|^2, |g^\flat(x)|^p)$$



for all points  $x \in \mathbb{R}^d$ . Therefore

$$1 - O(\varepsilon + \delta) \geq \int |g^\sharp|^p + c_p \int \min(|g^\sharp|^{p-2}|g^\flat|^2, |g^\flat|^p).$$

Let  $\tilde{\varepsilon} > 0$  be another small quantity. Then

$$1 - O(\varepsilon + \delta) \geq \|g^\sharp\|_p^p + c\tilde{\varepsilon}^{2-p} \int_{|g^\flat(x)| > \tilde{\varepsilon}|g^\sharp(x)|} |g^\flat(x)|^p dx$$

and consequently by (9.2),

$$\tilde{\varepsilon}^{2-p} \int_{|g^\flat(x)| > \tilde{\varepsilon}|g^\sharp(x)|} |g^\flat(x)|^p dx = O(\varepsilon + \delta)$$

and hence

$$\|g^\flat\|_p^p \leq \tilde{\varepsilon}^p \|g^\sharp\|_p^p + C\tilde{\varepsilon}^{p-2}(\varepsilon + \delta) \leq C\tilde{\varepsilon}^p + C\tilde{\varepsilon}^{p-2}(\varepsilon + \delta).$$

By choosing first  $\varepsilon, \delta$  to be small, then  $\tilde{\varepsilon} = (\varepsilon + \delta)^{1/2(2-p)}$ , we arrive at the desired bound.  $\square$

It is natural to ask whether the set of all extremizers of (1.2) satisfying  $\|f\|_p = 1$  is unique modulo the action of  $\mathcal{G}_d$ . We can offer only the following weak substitute.

**Corollary 9.4.** *Let  $\{F_\nu\}$  be any sequence of extremizers of (1.2) which satisfy  $\|F_\nu\|_p = 1$ . Then there exist a subsequence  $\{F_{\nu_k}\}$  and a sequence of symmetries  $\phi_k \in \mathcal{G}_d$  such that the sequence  $\{\phi_k^* F_{\nu_k}\}$  is convergent in  $L^p(\mathbb{R}^d)$  norm.*

This follows directly from Lemma 9.3.

## 10. ON THE SYMMETRY GROUPS

To list all elements of  $\mathcal{G}_{d,d}$ , we employ coordinates  $(x, y) = (x', x_d; y', y_d)$  with  $x', y' \in \mathbb{R}^{d-1}$  and  $x_d, y_d \in \mathbb{R}^1$ .  $\mathcal{G}_{d,d}$  is a subset of the set of all elements  $\Phi$  of  $\text{Diff}(\mathbb{R}^d) \times \text{Diff}(\mathbb{R}^d)$  of the form

$$(10.1) \quad (x', x_d; y', y_d) \mapsto (Lx' + u, tx_d + a + x' \cdot v + Q(x'); \tilde{L}y' + \tilde{u}, ty_d + \tilde{a} + y' \cdot \tilde{v} + \tilde{Q}(y')),$$

where  $L, \tilde{L}$  are linear endomorphisms of  $\mathbb{R}^{d-1}$ ,  $v, \tilde{v} \in \mathbb{R}^{d-1}$ ,  $0 \neq t \in \mathbb{R}$ ,  $a, \tilde{a} \in \mathbb{R}$ , and  $Q, \tilde{Q} : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$  are homogeneous quadratic polynomials. The proof of the next lemma is a straightforward verification, left to the reader.

**Lemma 10.1.** *A mapping  $\Phi$  of the form (10.1) belongs to  $\mathcal{G}_{d,d}$  if and only if all of the following equations hold:*

$$-a - \tilde{a} + 2u \cdot \tilde{u} - |u|^2 - |\tilde{u}|^2 = 0$$

$$\tilde{L}^* L = tI \text{ for some } t \in \mathbb{R} \setminus \{0\}$$

$$v + 2L^*(\tilde{u} - u) = 0$$

$$\tilde{v} + 2\tilde{L}^*(\tilde{u} - u) = 0$$

$$Q(x') = |Lx'|^2 - t|x'|^2$$

$$\tilde{Q}(y') = |\tilde{L}y'|^2 - t|y'|^2.$$

Given any  $L, u, t, a, v$  there exist  $Q, \tilde{Q}, \tilde{L}, \tilde{u}, \tilde{a}, \tilde{v}$  satisfying all these equations, and these quantities are uniquely determined by  $L, u, t, a, v$ .

**Definition 10.2.** A collection of  $d$  points  $x_j \in \mathbb{R}^d$  is said to lie in general position if, writing  $x_j = (x_{j,1}, \dots, x_{j,d})$ , the  $d \times d$  matrix

$$\begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d-1} & 1 \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d-1} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{d,1} & x_{d,2} & \cdots & x_{d,d-1} & 1 \end{pmatrix}$$

is nonsingular.

This definition is coordinate-dependent. The  $d$ -th coordinates of the points  $x_j$  do not enter into consideration.

**Lemma 10.3.** *The action of  $\mathcal{G}_d$  on  $\mathbb{R}^d$  is generically  $d$ -fold transitive. That is, for any two sets  $\{x_j : 1 \leq j \leq d\}$  and  $\{y_j : 1 \leq j \leq d\}$  of  $d$  points in general position in  $\mathbb{R}^d$ , there exists  $\phi \in \mathcal{G}_d$  satisfying*

$$\phi(x_j) = y_j \text{ for all } 1 \leq j \leq d.$$

*Proof.* There exists a unique affine endomorphism  $x' \mapsto Lx' + u$  of  $\mathbb{R}^{d-1}$  which maps  $x'_j$  to  $y'_j$  for all  $1 \leq j \leq d$ . Fix any  $0 \neq t \in \mathbb{R}$ . Define  $Q(y') = |L(y')|^2 - t|y'|^2$  for  $x' \in \mathbb{R}^{d-1}$ . Let  $a \in \mathbb{R}$  and  $v \in \mathbb{R}^{d-1}$  be parameters to be specified. Consider the diffeomorphism  $\phi \in \mathcal{G}_d$  defined by  $\phi(x', x_d) = (Lx' + u, tx_d + a + v \cdot x' + Q(x'))$ .

We merely need to choose these parameters to satisfy the equations

$$tx_{j,d} + a + v \cdot x'_j + Q(x'_j) = y_{j,d} \text{ for all } 1 \leq j \leq d,$$

which is to say,

$$v \cdot x'_j + a = y_{j,d} - |L(x'_j)|^2 - t(x_{j,d} - |x'_j|^2).$$

Since  $t$  has already been specified, this is a system of  $d$  inhomogeneous linear equations for  $(v, a) \in \mathbb{R}^{(d-1)+1}$ . The assumption that  $\{x_j : 1 \leq j \leq d\}$  is in general position means precisely that this system has full rank  $d$ . Thus for any  $t$ , there exists a (unique) solution.  $\square$

The proof also demonstrates that  $\phi$  is uniquely determined, up to the “vertical” scaling factor  $t$ .

## 11. AFFINE SURFACE MEASURE

Our definition of  $\mathbf{T}$  uses the measure  $dt$ , rather than the surface measure on the paraboloid induced from its inclusion into  $\mathbb{R}^d$ . This measure  $dt$  is entirely natural from a geometric viewpoint, which we now explain.

First, we recall the definition and properties of affine *arclength* measure. This is a measure on a subinterval  $I \subset \mathbb{R}$ , associated to any sufficiently smooth mapping  $\gamma : I \rightarrow \mathbb{R}^d$ , as follows: Define

$$L_\gamma(t) = |\det(\gamma'(t), \gamma^{(2)}(t), \dots, \gamma^{(d)}(t))|$$

where  $\gamma^{(k)}(t) = d^k \gamma(t)/dt^k$ . Then the measure  $\sigma_\gamma(I)$  is defined to be

$$(11.1) \quad \sigma_\gamma(I) = \int_I |L_\gamma(t)|^{2/d(d+1)} dt.$$

See [8] for some discussion.

Affine arclength measure  $\sigma_\gamma$  enjoys two natural invariances, which provide its *raison d'être*:

- (i) For any  $A \in Gl(\mathbb{R}, d)$ ,  $\sigma_{A \circ \gamma}(I) = |\det(A)|^{2/d(d+1)} \sigma_\gamma(I)$ .
- (ii) For any injective  $C^d$  mapping  $\phi : I \rightarrow \mathbb{R}$ ,  $\sigma_{\gamma \circ \phi}(I) = \sigma_\gamma(\phi(I))$ .

Both identities are easily verified.

There is a natural analogue in the codimension one case, which we call affine surface measure. Let  $F$  be any  $C^2$  mapping from an open subset of  $\mathbb{R}^{d-1}$  to  $\mathbb{R}^d$ . Write  $F(t) = (F_1(t), \dots, F_d(t))$ . For each pair of indices  $i, j \in \{1, 2, \dots, d-1\}$ , form

$$F_{i,j}(t) = \det \begin{pmatrix} \frac{\partial F_1(t)}{\partial t_1} & \dots & \frac{\partial F_1(t)}{\partial t_{d-1}} & \frac{\partial^2 F_1(t)}{\partial t_i \partial t_j} \\ \frac{\partial F_2(t)}{\partial t_1} & \dots & \frac{\partial F_2(t)}{\partial t_{d-1}} & \frac{\partial^2 F_2(t)}{\partial t_i \partial t_j} \\ \vdots & \dots & \vdots & \vdots \\ \frac{\partial F_d(t)}{\partial t_1} & \dots & \frac{\partial F_d(t)}{\partial t_{d-1}} & \frac{\partial^2 F_d(t)}{\partial t_i \partial t_j} \end{pmatrix}.$$

From these quantities  $F_{i,j}$  we form a  $(d-1) \times (d-1)$  matrix, and consider its determinant

$$\mathcal{L}_F(t) = \det (F_{i,j}(t))_{i,j=1}^{d-1}.$$

For any Borel set  $U \subset \mathbb{R}^{d-1}$  contained in the domain of  $F$ , define the affine surface measure of  $U$  to be

$$\sigma_F(U) = \int_U |\mathcal{L}_F(t)|^{1/(d+1)} dt.$$

Affine surface measure enjoys corresponding invariances.

**Proposition 11.1.** *Let  $A \in Gl(\mathbb{R}, d)$  and let  $\phi : V \rightarrow U$  be a  $C^2$  diffeomorphism of an open set  $V \subset \mathbb{R}^{d-1}$  with a subset  $U$  of the domain of  $F$ . Then*

$$(11.2) \quad \sigma_{A \circ F}(U) = |\det(A)|^{(d-1)/(d+1)} \sigma_F(U)$$

$$(11.3) \quad \sigma_{F \circ \phi}(V) = \sigma_F(\phi(V)).$$

In the case where  $F(t) = (t_1, \dots, t_{d-1}, |t|^2)$ , affine surface measure is a constant multiple of Lebesgue measure on  $\mathbb{R}^{d-1}$ , which is the measure used in our definition of **T**. Thus our use of this measure, rather than of surface measure on the paraboloid, has a natural geometric context.

*Proof of Proposition 11.1.* Property (11.2) follows directly from the definition; passing from  $F$  to  $A \circ F$  multiplies each  $F_{i,j}$  by  $\det(A)$ . To derive (11.3), note first that if  $G = F \circ \phi$ , then for any indices  $i, j$ , the determinant  $G_{i,j}(t)$  depends only on the Jacobian matrix  $D\phi(t)$  and on the first and second partial derivatives of  $F$  at  $\phi(t)$ . Indeed, the chain rule produces an undesired contribution involving first derivatives

of  $F$  and quadratic expressions involving  $D\phi$  in the rightmost column, but this contribution is a linear combination of the first  $d - 1$  columns and hence can be eliminated via row operations.

Thus matters reduce to the case where  $\phi$  is linear, and it is no loss of generality to suppose that  $t = 0$  and  $F(0) = 0$ . We claim then that

$$(11.4) \quad |\mathcal{L}_{F \circ \phi}(0)| = |\det(\phi)|^{d+1} |\mathcal{L}_F(0)|.$$

Taking into account both the formula  $d\tau = |\det(D\phi)| dt$ , where  $\tau = \phi(t)$ , and the exponent  $1/(d + 1)$  in the definition of  $\sigma_{F \circ \phi}$ , would then yield (11.3).

To derive (11.4), the invariance (11.2) already established can be used to reduce matters to the case where  $F$  takes the form  $F(t) = (t_1, t_2, \dots, t_{d-1}, f(t))$  where  $Df(0) = 0$ . Then

$$(F \circ \phi)_{i,j}(0) = \det(\phi) \cdot \sum_{k,l} \partial_{i,j}^2 f(0) \phi_{k,i} \phi_{l,j}$$

where  $\partial_{i,j}^2 f$  denotes the second partial derivatives of  $f$ , and  $\phi = (\phi_{k,i})_{k,i=1}^{d-1}$ . Thus

$$\mathcal{L}_{F \circ \phi}(0) = \det(\phi)^{d-1} \cdot \det \left( \sum_{k,l} \partial_{i,j}^2 f(0) \phi_{k,i} \phi_{l,j} \right)_{i,j=1}^{d-1}.$$

This last matrix is simply the composition  $\phi^* \circ \partial^2 f(0) \circ \phi$  where  $\phi^*$  denotes the transpose of  $\phi$ . Therefore

$$\mathcal{L}_{F \circ \phi}(0) = \det(\phi)^{d-1} \cdot \det(\phi)^2 \det(\partial^2 f(0)) = \det(\phi)^{d+1} \mathcal{L}_F(0),$$

which is (11.4). □

Inequalities for operators defined by convolution with (the push-forward onto a curve of) affine arclength are studied in [7] and [11]. See the former paper for references to other related works. One may ask whether there are analogues, for hypersurfaces equipped with affine surface measure, of the results of [7], [11] for curves equipped with affine arclength measure.

## REFERENCES

- [1] M. Christ, *Estimates for the  $k$ -plane transform*, Indiana Univ. Math. J. 33 (1984), no. 6, 891–910
- [2] ———, *Convolution, combinatorics, and curvature: A case study*, Internat. Math. Res. Notices 1998, no. 19, 1033–1048.
- [3] ———, *Quasixtrems for a Radon-like transform*, preprint.
- [4] ———, *Extremizers of a Radon transform inequality*, preprint.
- [5] M. Christ and S. Shao, *Existence of Extremals for a Fourier Restriction Inequality*, preprint, arXiv:1006.4319. To appear in Analysis and Partial Differential Equations
- [6] M. Christ and Q. Xue, *Smoothness of extremizers of a convolution inequality*, preprint, arXiv:1012.5458
- [7] S. Dendrinos, N. Laghi, and J. Wright, *Universal  $L^p$  improving for averages over polynomial curves in low dimensions*, J. Funct. Anal. 257 (2009), no. 5, 1355–1378.
- [8] H. W. Guggenheimer, *Differential geometry*. McGraw-Hill, New York, 1963.
- [9] B. Stovall, *Endpoint bounds for a generalized Radon transform*, J. Lond. Math. Soc. (2) 80 (2009), no. 2, 357–374.
- [10] ———, *Quasi-extremals for convolution with surface measure on the sphere*, Illinois J. Math. 53 (2009), no. 2, 391–412.

- [11] ———, *Endpoint  $L^p \rightarrow L^q$  bounds for integration along certain polynomial curves*, J. Funct. Anal. 259 (2010), no. 12, 32053229.
- [12] E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton Mathematical Series, no. 32. Princeton University Press, Princeton, N.J., 1971.

MICHAEL CHRIST, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, BERKELEY,  
CA 94720-3840, USA

*E-mail address:* mchrist@math.berkeley.edu